11

Handling Omitted Variable Bias in Multilevel Models: Model Specification Tests and Robust Estimation

Jee–Seon Kim and Chris M. Swoboda
Department of Educational Psychology,
University of Wisconsin at Madison, Madison, Wisconsin

11.1 INTRODUCTION

Multilevel models allow researchers to examine hypothesized relationships across different units of analysis in a statistically appropriate way, thus permitting more accurate modeling of complex systems. At the same time, the complexity of multilevel models introduces other challenges in statistical modeling, as many assumptions are needed (Goldstein, 2003; Hox, 2002; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). In particular, there are multiple random effects in multilevel models and it is assumed that all predictors in the model are uncorrelated with all of the random effects. Standard estimation methods for multilevel models such as full information maximum likelihood (FIML), restricted (or residual) maximum likelihood (REML), generalized least squares (GLS), empirical Bayes, and fully Bayesian estimators all assume the independence of predictors from random effects and would yield biased estimates if the assumption is violated.

However, independence between predictors and random effects is prone to be violated in practice. There are three common forms of bias due to correlated effects (Kim & Frees, 2007). First, unobserved effects can lead to *omitted variable bias*. Second, predictors might be measured imprecisely and result in *measurement error* or *error-in-variable bias*. Third, some predictors may not only cause but also be influenced by the outcome variable (two-way causality), yielding *simultaneity bias*. Among these, this chapter focuses on handling correlated effects due to omitted variables, which is a common thread in most observational and quasi-experimental studies in the social and behavioral sciences.

Whereas model specification tests have been one of the most important areas of research in econometrics for decades (Frees, 2004; Hausman, 1978; Wooldridge, 2002), specification issues have been often overlooked in multilevel analysis. We argue that concerns for omitted variables and other specification issues should be routine in multilevel analysis. An omitted variable at one level may yield severe bias at all levels in terms of regression coefficients as well as variance components (Kim, 2009).

This chapter provides a tutorial for the recent statistical developments in Kim and Frees (2006, 2007), which introduced a set of statistical tools for testing the severity of omitted variable bias and for obtaining robust estimators in the presence of omitted variable effects. While these two original articles are more technical, this chapter provides a more conceptual review of the methodology with less emphasis on mathematics. An application of the methods is illustrated with a well-known data set, the National Education Longitudinal Study of 1988 (NELS:88). For details on the procedures and longitudinal data examples, we refer to the original papers.

The rest of the chapter is organized as follows: We define fundamental concepts such as endogeneity and exogeneity in linear models and provide a review of econometric treatment of omitted variables in Section 11.2. Sections 11.3 and 11.4 present omitted variable tests and the *generalized method of moments* (GMM) estimation technique, respectively. Section 11.5 applies the methodologies of the previous two sections to NELS:88. The last section provides a summary and ends with a discussion of related topics in various disciplines.

11.2 BACKGROUNDS

11.2.1 Endogeneity and Omitted Variable Bias

The problem of *endogeneity* in a regression model occurs when a predictor is correlated with the error term in the model. Endogeneity is defined similarly in multilevel modeling but, unlike a regression model, there exist multiple random components in a multilevel model and thus more opportunities for endogeneity to occur.

We consider a linear model as:

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

where $\mathbf{X}$ is the collection of predictors across the levels and $\boldsymbol{\delta}$ is the collection of all random components. If a predictor is correlated with $\boldsymbol{\delta}$, it is an *endogenous* variable. If not, it is an *exogenous* variable. Thus, the *exogeneity* assumption implies that all predictors are uncorrelated with all random components in the model.

As this chapter focuses on linear models, we express the condition for exogeneity as:

$$\mathrm{E}(\mathbf{X}\boldsymbol{\delta}) = \mathbf{0}. \qquad (11.1)$$

Variables that do not satisfy this condition are said to be endogenous. Sometimes, a more restrictive assumption $\mathrm{E}(\boldsymbol{\delta}|\mathbf{X}) = \mathbf{0}$ is also used as the exogeneity condition.

To illustrate the problem of omitted variables, consider a "true" model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\gamma} + \boldsymbol{\delta}, \qquad (11.2)$$

where $\mathbf{U}$ are unobserved predictors that affect the outcome. Since $\mathbf{U}$ is unobserved—hence omitted in the analysis—the "fitted" model is

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

where $\tilde{\boldsymbol{\delta}} = \mathbf{U}\boldsymbol{\gamma} + \boldsymbol{\delta}$. The expected value of the least squares estimates for the regression coefficients associated with **X** can be shown to be $\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}\boldsymbol{\gamma}$. Unless either $\mathbf{X}'\mathbf{U} = \mathbf{0}$ (observed and unobserved predictors are uncorrelated) or $\boldsymbol{\gamma} = 0$ (unobserved predictors do not affect the outcome), $E(\mathbf{X}\tilde{\boldsymbol{\delta}}) \neq 0$ and the least squares estimator of $\boldsymbol{\beta}$ is biased and inconsistent.

In multilevel models, GLS estimators, ML estimators, and empirical Bayes estimators often provide indistinguishable solutions. GLS solutions are sometimes used as starting values for other estimators for complex multilevel models or a large data set. Importantly, all of these estimators rely on the assumption that predictors are uncorrelated with random components. However, this exogeneity assumption is prone to be violated in most observational and quasi-experimental studies, where researchers do not have the ability to control for or collect all the right variables.

The endogeneity problem has been studied extensively in econometrics, mainly in the context of panel data analysis (Frees, 2004; Frees & Kim, 2008; Hsiao, 2003), and statistical tests for omitted variable effects such as the Hausman test (1978) have been steadily used for the past 30 years. In the following section, we review the econometric treatment of omitted variable bias and explain why the methodology for panel data models is not appropriate for more complex multilevel models.

11.2.2 Econometric Treatment and Its Limitations

A panel data model that includes omitted variables can be written as

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\,\boldsymbol{\beta} + u_i + \varepsilon_{it}, \qquad (11.3)$$

where individual $i = 1,\ldots,n$ is observed over time points $t = 1,\ldots,T_i$. Equation 11.3 includes the outcome variable y_{it} (test score, for example), disturbance term ε_{it}, predictors $\mathbf{x}_{it}$, and coefficient vector $\boldsymbol{\beta}$. The model also contains a latent intercept variable α_i that is constant over time. This latent variable induces a correlation among individual responses over time and serves as a proxy for unobserved time-constant characteristics, such as "ability," that are uncorrelated with the predictors. Without the omitted variable u_i, the model in Equation 11.3 is a two-level random intercept model.

Unlike α_i, u_i may be correlated with one or more of the predictors in $\mathbf{x}_{it}$. Thus, this variable may create bias in the estimates of $\boldsymbol{\beta}$. To mitigate the effects of u_i, one can apply a fixed effects (FE) transformation, sweeping out the time-constant omitted variable. Here, the phrase "sweeping out" refers to the fact that the FE transformation in Equation 11.4 below would remove all time-constant variables. To see the impact of the FE transformation strategy, take averages over time in Equation 11.3 to get

$$\bar{y}_i = \alpha_i + \mathbf{x}'_i\,\boldsymbol{\beta} + u_i + \bar{\varepsilon}_i.$$

Subtracting this from Equation 11.3 yields the transformed model equation:

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\,\boldsymbol{\beta} + \varepsilon_{it} - \bar{\varepsilon}_i. \quad (11.4)$$

Then the least squares estimator of β is

$$\mathbf{b}_{\mathrm{FE}} = \left(\sum_i \sum_t (\mathbf{x}_{it} - \mathbf{x}_i)'(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\right)^{-1} \times \left(\sum_i \sum_t (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'(y_{it} - \bar{y}_i)\right). \qquad (11.5)$$

This estimator is unbiased even in the presence of the time-constant omitted variable u_i. We denote this estimator as a fixed effects (FE) estimator $\mathbf{b}_{\text{FE}}$. One important strength of the FE estimator is that there are relatively few assumptions needed, compared to alternative procedures. For example, instrumental variable estimation requires the analyst to identify a proxy for the omitted variable. Similarly, simultaneous equations modeling requires specifying a model for the latent, unobserved, omitted variables. Although these alternatives are certainly appropriate in many circumstances, they do require additional (and sometimes unavailable) knowledge from the analyst.

Another important advantage is that these procedures are easy to implement. Generally, one can implement the calculations using standard statistical packages while only specifying certain "fixed effects" nuisance parameters, or unwanted "dummy variables." Note that we do not actually introduce these extra unwanted parameters into the model; the estimates are simply a by-product of the estimation procedures when Equation 11.5 is calculated by a computer program that is not specifically designed for the FE estimator.

Equation 11.5 also underscores a major limitation of FE estimation. Note that this estimator only provides estimates for variables that vary by time; if the jth predictor of $\mathbf{x}_{it}$ is constant over time, then the jth row of $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$ is identically zero so that $\boldsymbol{\beta}$ is not estimable. This means that while the FE approach removes undesirable u_i, it also sweeps out potentially important $\bar{\mathbf{x}}_i$ such as variables related to family characteristics, teacher qualities, and school environments. This loss of information would be a critical drawback in many applications.

Therefore, if bias due to omitted variables is not significant, one may prefer other options that do not lose information. To evaluate the size of omitted variable bias, we again consider the model in Equation 11.3 but now assume that there are no omitted variables so that $u_i = 0$. Without u_i, predictors are uncorrelated with random effects in the model and one can estimate $\boldsymbol{\beta}$ using standard procedures discussed earlier, including REML, FIML, GLS, and Bayesian estimators. We refer to a resulting estimator using the independence assumption between predictors and random effects as a random effects (RE) estimator, denoted as $\mathbf{b}_{\text{RE}}$. Note that $\mathbf{b}_{\text{RE}}$ would be biased in the presence of omitted variables ($u_i \neq 0$).

The severity of omitted variable bias for a given data set can be examined by comparing FE with RE estimators, as the former is robust to the presence of u_i and the latter is not. Hausman (1978) presented a test for examining the effect of omitted variables by measuring the distance between vectors $\mathbf{b}_{\text{FE}}$ and $\mathbf{b}_{\text{RE}}$:

$$\chi^2_{\text{Hausman}} = (\mathbf{b}_{\text{FE}} - \mathbf{b}_{\text{RE}})' (Var\,\mathbf{b}_{\text{FE}} - Var\,\mathbf{b}_{\text{RE}})^{-1} \times (\mathbf{b}_{\text{FE}} - \mathbf{b}_{\text{RE}}). \tag{11.6}$$

This test statistic follows a chi-square distribution under the null hypothesis of no omitted variables, with degrees of freedom equal to the number of parameters in $\mathbf{b}_{\text{FE}}$. Under the null hypothesis that there exist no omitted variables, both $\mathbf{b}_{\text{FE}}$ and $\mathbf{b}_{\text{RE}}$ are unbiased. In contrast, only $\mathbf{b}_{\text{FE}}$ is unbiased if the null hypothesis is not true. Thus, when the hypothesis is retained, the analyst would choose $\mathbf{b}_{\text{RE}}$ without the loss of information in $\mathbf{b}_{\text{FE}}$. On the other hand, if the distance between the two estimators is statistically significant, this implies that the random effects estimator is biased due to the presence of omitted variable effects, and the robust fixed effects estimator should be used despite its limitations.

The Hausman test is an effective tool for panel data models, but its applicability is limited for multilevel models, because the test was developed for examining omitted variable effects at the second level in two level models, not for three- or higher-level models. If there exist omitted variable effects at a lower level (e.g., teacher level), applying the Hausman test at a higher level (e.g., school level) would yield fallacious results, as both estimators in the test would be biased. Simulation studies have shown that the level-3 (school) fixed effects estimator could be as severely biased as the random effects estimator when there exist level-2 (teacher) omitted variable effects (Kim & Frees, 2006). This is not a problem of the Hausman test itself, but rather is an improper use of the test for which it is not designed.

Therefore, while we recommend the Hausman test for a panel data model, which is mathematically equivalent to a two-level random intercept model, it should be noted that the test is not appropriate for examining more complex models with more than two levels and/or random slopes. Test results can be highly misleading in more general multilevel modeling contexts. On the one hand, retaining the null hypothesis may not imply the absence of omitted variable effects but may reflect that both the fixed and random effects estimators are biased in a similar way. On the other hand, rejecting the null hypothesis may suggest that the fixed and random effects estimators are biased in a different way.

In the rest of the chapter, we explain recent methodology for extending the idea of the Hausman test to be applicable in more general multilevel models. Extensions involve multiple hypotheses and corresponding tests to determine the location and severity of omitted variable effects. Moreover, we present an alternative robust estimation approach that overcomes the loss of information problem of the FE estimator. The methodology and its implications are presented in the context of a model for mathematics achievement in NELS:88. This chapter ends with a discussion of related topics in various disciplines.

11.3 OMITTED VARIABLES IN MULTILEVEL MODELS

A key to understanding the methodology for handling omitted variable bias is to recognize that when we have only one set of estimates, we do not know if those estimates are biased or not. To diagnose bias, we need to consider multiple sets of estimates obtained from different estimators, ideally where some are more robust than others. This section presents notation for a three-level model, defines two FE estimators and the RE estimator in the three-level model, and introduces statistical tests to compare these estimators.

11.3.1 Multilevel Models with Unobserved Variables

Consider three levels of nesting, where the subscript s identifies a school, the subscript t identifies a teacher belonging to school s, and the subscript p denotes a pupil belonging to school s and teacher t. The level-1 model is then written as:

$$y_{stp} = \mathbf{Z}_{stp}^{(1)}\boldsymbol{\beta}_{st}^{(1)} + \mathbf{X}_{stp}^{(1)}\boldsymbol{\beta}_1 + \varepsilon_{stp}^{(1)},$$

where y_{stp} denotes the response variable. Predictors $\mathbf{Z}_{stp}^{(1)}$ and $\mathbf{X}_{stp}^{(1)}$ may be related to the pupil, teacher, or school. Parameters that are constant appear in the $\boldsymbol{\beta}_1$ vector and so we interpret $\mathbf{X}_{stp}^{(1)}\boldsymbol{\beta}_1$ to be part of the "fixed effects" portion of the model. The term $\boldsymbol{\beta}_{st}^{(1)}$ captures latent, unobserved characteristics that are school and teacher specific. We wish to allow for, and test, the possibility that these latent characteristics are related to predictors $\mathbf{Z}_{stp}^{(1)}$ and $\mathbf{X}_{stp}^{(1)}$. For identification purposes, we adopt the usual convention and assume that the disturbance term, $\varepsilon_{stp}^{(1)}$, is independent of the other right-hand variables, $\mathbf{Z}_{stp}^{(1)}$, $\mathbf{X}_{stp}^{(1)}$, and $\boldsymbol{\beta}_1$.

The level-2 model describes the variability at the teacher level and is written as

$$\boldsymbol{\beta}_{st}^{(1)} = \mathbf{Z}_{st}^{(2)}\boldsymbol{\beta}_s^{(2)} + \mathbf{X}_{st}^{(2)}\boldsymbol{\beta}_2 + \varepsilon_{st}^{(2)}.$$

Analogous to the level-1 model, the predictors $\mathbf{Z}_{st}^{(2)}$ and $\mathbf{X}_{st}^{(2)}$ relate to the teacher or school. Constant parameters appear in the $\boldsymbol{\beta}_2$ vector and so we interpret $\mathbf{X}_{st}^{(2)}\boldsymbol{\beta}_2$ to be the fixed effects at the teacher level. The term $\boldsymbol{\beta}_s^{(2)}$ captures latent, unobserved characteristics that are school specific; these latent characteristics may be related to predictors $\mathbf{Z}_{st}^{(2)}$ and $\mathbf{X}_{st}^{(2)}$.

Finally, the level-3 model describes variability at the school level, and is written as

$$\boldsymbol{\beta}_s^{(2)} = \mathbf{X}_s^{(3)}\boldsymbol{\beta}_3 + \boldsymbol{\varepsilon}_s^{(3)}.$$

The variables $\mathbf{X}_s^{(3)}$ may depend on the school. We let $\boldsymbol{\varepsilon}_s^{(3)}$ represent other unobserved characteristics of the school that are not explained by the fixed effects portion, $\mathbf{X}_s^{(3)}\boldsymbol{\beta}_3$.

It is well known that a multilevel model may be written as a linear mixed-effects model for, among other reasons, parameter estimation purposes. For fitting multilevel models as linear mixed-effects models, Singer (1998) provides an insightful overview using SAS PROC MIXED examples. Combining the separate models for the three levels above, the multilevel model can be expressed as a linear-mixed effects model:

$$y_{stp} = \mathbf{Z}_{stp}^{(1)}\mathbf{Z}_{st}^{(2)}\left(\mathbf{X}_s^{(3)}\boldsymbol{\beta}_3 + \boldsymbol{\varepsilon}_s^{(3)}\right) + \mathbf{Z}_{stp}^{(1)}\left(\mathbf{X}_{st}^{(2)}\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_{st}^{(2)}\right) + \mathbf{X}_{stp}^{(1)}\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_{stp}^{(1)}. \quad (11.7)$$

Next, define $\mathbf{Z}_{2,stp} = \mathbf{Z}_{stp}^{(1)}$ and $\mathbf{Z}_{3,stp} = \mathbf{Z}_{stp}^{(1)}\mathbf{Z}_{st}^{(2)}$. With this notation, we may summarize all random component terms as $\delta_{stp} = \mathbf{Z}_{3,stp}\,\varepsilon_s^{(3)} + \mathbf{Z}_{2,stp}\,\varepsilon_{st}^{(2)} + \varepsilon_{stp}^{(1)}$. Further, define the $K \times 1$ vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3')$ and the $1 \times K$ vector $\mathbf{X}_{stp} = \left(\mathbf{X}_{stp}^{(1)} : \mathbf{Z}_{2,stp}\mathbf{X}_{st}^{(2)} : \mathbf{Z}_{3,stp}\mathbf{X}_s^{(3)}\right)$. With this notation and stacking, we may express the model as

$$\mathbf{y}_s = \mathbf{X}_s\boldsymbol{\beta} + \boldsymbol{\delta}_s, \quad (11.8)$$

which emphasizes that schools are independent units in the model. The exogeneity assumption would be violated if some of $\mathbf{X}_s$ are correlated with $\boldsymbol{\delta}_s$.

It is useful to consider a special and common case of this general three-level model where there are no random slopes, but the intercepts are associated with random coefficients. Using the above notation, this implies $\mathbf{Z}_{stp}^{(1)} = 1$ and $\mathbf{Z}_{st}^{(2)} = \mathbf{1}$. Multilevel models of this kind are referred to as *random intercept models*. We can write the three-level random intercept model as a single equation of the form:

$$y_{stp} = \mathbf{X}_s^{(3)}\boldsymbol{\beta}_3 + \mathbf{X}_{st}^{(2)}\boldsymbol{\beta}_2 + \mathbf{X}_{stp}^{(1)}\boldsymbol{\beta}_1 + \varepsilon_s^{(3)} + \varepsilon_{st}^{(2)} + \varepsilon_{stp}^{(1)}. \quad (11.9)$$

To illustrate omitted variable problems in multilevel models, we denote unobserved effects of schools and teachers as $\mathbf{u}_s^{(3)}$ and $\mathbf{u}_{st}^{(2)}$, respectively, and add them to Equation 11.9:

$$y_{stp} = \mathbf{X}_s^{(3)}\boldsymbol{\beta}_3 + \mathbf{X}_{st}^{(2)}\boldsymbol{\beta}_2 + \mathbf{X}_{stp}^{(1)}\boldsymbol{\beta}_1 + \varepsilon_s^{(3)} + \varepsilon_{st}^{(2)} + \varepsilon_{stp}^{(1)} + \mathbf{u}_s^{(3)} + \mathbf{u}_{st}^{(2)}. \tag{11.10}$$

The latent intercept variables $\varepsilon_s^{(3)}$ and $\varepsilon_{st}^{(2)}$ are uncorrelated with the predictors. By contrast, $\mathbf{u}_s^{(3)}$ and $\mathbf{u}_{st}^{(2)}$ may be correlated with one or more of the predictors in the model and thus their omission may create bias in the estimates of $\boldsymbol{\beta}$. In the next section, we present omitted variable tests to examine bias due to $\mathbf{u}_s^{(3)}$ and $\mathbf{u}_{st}^{(2)}$ in a given data set.

11.3.2 Different Estimators

As mentioned earlier, commonly used multilevel model estimators such as REML, FIML, GLS, and Bayesian estimators assume that $\mathbf{u}_s^{(3)} = \mathbf{u}_{st}^{(2)} = 0$ in Equation 11.10, and each of these estimators can be considered as a random effects (RE) estimator, $\mathbf{b}_{\text{RE}}$, which provides an unbiased solution in the absence of omitted variables but yields biased estimates of model parameters if the assumption is violated. To examine the degree of bias in b_{RE}, we also considered the fixed teacher effects estimator $\mathbf{b}_{\text{FEt}}$, which is robust against the presence of both $\mathbf{u}_s^{(3)}$ and $\mathbf{u}_{st}^{(2)}$, and the fixed school effects estimator $\mathbf{b}_{\text{FEs}}$, which is robust against the presence of $\mathbf{u}_s^{(3)}$ but not $\mathbf{u}_{st}^{(2)}$.

The two FE estimators, $\mathbf{b}_{\text{FEt}}$ and $\mathbf{b}_{\text{FEs}}$, can be obtained by treating the teacher and school identification variables as discrete variables. This may take some time and the output can be long (unless suppressed), if the number of units is large, because the program would estimate nuisance parameters for the dummy variables assigned to the entities. In Section 11.4, we present a direct approach to obtain $\mathbf{b}_{\text{FEt}}$ and $\mathbf{b}_{\text{FEs}}$ without the unnecessary dummy variables. Note that while $\mathbf{b}_{\text{RE}}$ provides estimates at all three levels, $\mathbf{b}_{\text{FEt}}$ only provides estimates for the level-1 variables while $\mathbf{b}_{\text{FEs}}$ provides estimates for the level-1 and level-2 variables. Because of this severe loss of information, one may not choose the FE estimators unless the bias in RE is statistically significant.

11.3.3 Three Types of Omitted Variable Tests

Kim and Frees (2006) presented three omitted variable tests for examining omitted variable bias in multilevel models. They can be recognized as multiple-level, intermediate-level, and highest-level tests, respectively. The names of the tests indicate the locations of potential omitted variables being tested. The *multiple-level* test can be viewed as an "omnibus" test for examining all potential omitted variable effects simultaneously across levels by comparing the most robust estimator to the most efficient estimator. If this test indicates that the most efficient but least robust estimator (i.e., $\mathbf{b}_{\text{RE}}$) is unbiased, no further test may be necessary and we can make inferences based on $\mathbf{b}_{\text{RE}}$.

In most applications, the analyst may have some ideas about omitted variables and want to test one or more levels separately. Omitted variable effects at lower levels can be tested by the *intermediate-level* test, regardless of omitted variable effects at higher levels. In a three-level school–teacher–pupil model, the omitted teacher effects can be tested regardless of omitted school effects. In a four-level model, omitted variable effects at the second and/or third level can be tested regardless of omitted variable effects at the fourth level using the intermediate-level test.

Finally, the *highest-level* test examines omitted variable effects at higher levels, assuming there exist no omitted variable effects at lower levels. In the school–teacher–pupil model, omitted school effects can be tested assuming there exist no omitted teacher effects. In a four-level model, omitted variable effects at the fourth level can be tested assuming there exist no omitted variable effects at the second and third levels. Also, omitted variable effects at the third and fourth level can be tested assuming there exist no omitted variable effects at the second level using the highest-level test.

All three tests can be defined by one test statistic:

$$\chi^2_{\text{OVT}} = (\mathbf{b}_{\text{robust}} - \mathbf{b}_{\text{efficient}})' \times (\text{Var}\,\mathbf{b}_{\text{robust}} - \text{Var}\,\mathbf{b}_{\text{efficient}})^{-1} \times (\mathbf{b}_{\text{robust}} - \mathbf{b}_{\text{efficient}}), \tag{11.11}$$

and a pair of robust and efficient estimators is determined with respect to the hypothesis being tested. For a three level model, the hypotheses and corresponding robust and efficient estimators for the three omitted variable tests can be summarized as follows:

Omitted Variable Test	Hypothesis	Robust Estimator	Efficient Estimator
1. Multiple-level test	$\mathbf{u}_s^{(3)} = \mathbf{u}_{st}^{(2)} = 0$	$\mathbf{b}_{\text{FEt}}$	$\mathbf{b}_{\text{RE}}$
2. Intermediate-level test	$\mathbf{u}_{st}^{(2)} = 0$	$\mathbf{b}_{\text{FEt}}$	$\mathbf{b}_{\text{FEs}}$
3. Highest-level test	$\mathbf{u}_s^{(3)} = 0$	$\mathbf{b}_{\text{FEs}}$	$\mathbf{b}_{\text{RE}}$

In regard to determining what tests would be appropriate for a given model, one may consider the following two properties. First, the omnibus multiple-level test would be recommended in general if omitted variable effects are of concern at all levels. When the multiple-level test is rejected, one can subsequently test each level using the other tests. However, if one suspects omitted variable effects at a particular level, the intermediate-level or highest-level test would be more powerful than the multiple-level test.

Second, it is important to note the asymmetry properties of the second and third tests; that is, while the intermediate-level test is valid regardless of omitted variable effects at higher levels, the highest-level test is only valid without omitted variable effects at lower levels. This occurs because the lower-level fixed effects estimators are robust against higher-level omitted variable effects, but the higher-level fixed effects estimators are not robust against lower-level omitted variable effects. Therefore, the analyst should either assume or test $\mathbf{u}_{st}^{(2)} = 0$ before testing $\mathbf{u}_s^{(3)} = 0$. The three omitted variable tests are shown at the top of Figure 11.1, which provides a flowchart for data analysis procedures for model specification tests and selecting the optimal estimator. Figure 11.1 will be revisited with an empirical example in Section 11.5.

11.4 GENERALIZED METHOD OF MOMENTS (GMM) INFERENCE

The generalized method of moments (GMM) is a general estimation method for statistical models. As its name indicates, it is

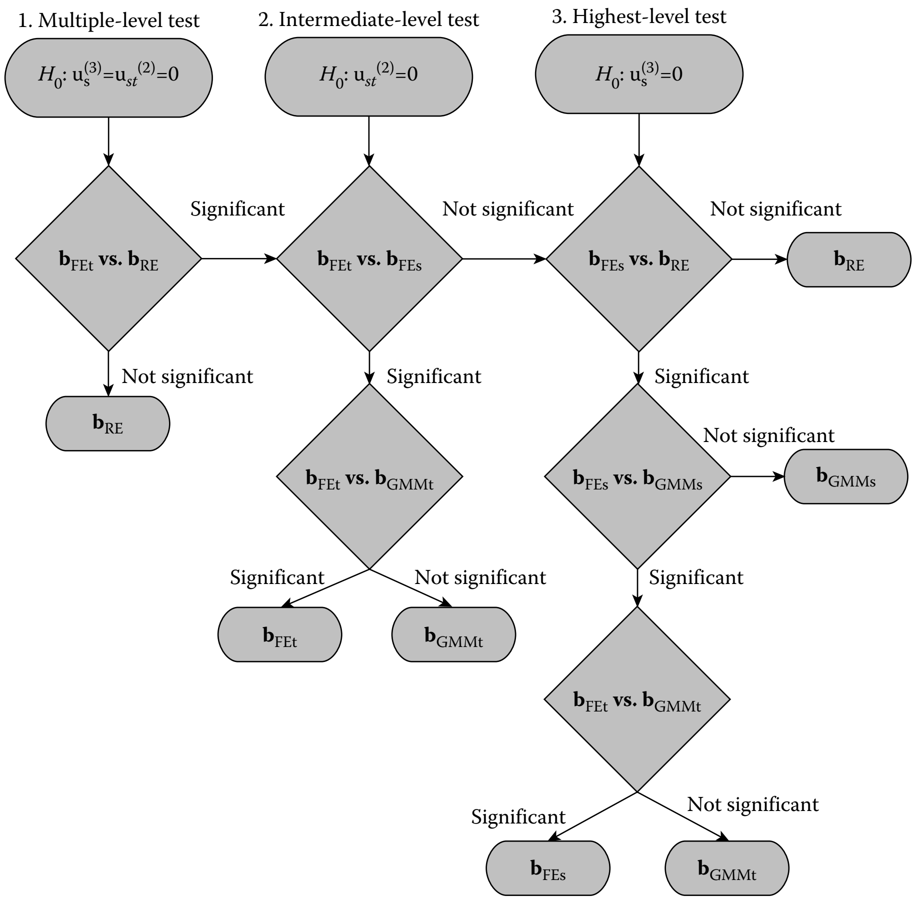

FIGURE 11.1
General guideline for conducting omitted variable tests and selecting the optimal estimator in a three-level model. One may start with any one of the three tests. However, Test 3 is valid when $\mathbf{U}_{st}^{(2)} = 0$. One should assume $\mathbf{U}_{st}^{(2)} = 0$ or conduct Test 2 beforehand.

a generalization of the method of moments (Hansen, 1982) and can also be viewed as an extension of instrumental variable (IV) methods. The area of GMM inference is more technical than other topics in this chapter and involves several statistical concepts such as instrument, projection, transformation, and generalized inverse. For this reason, we seek to convey the conceptual ideas of the methodology rather than its details. For GMM approaches in general, we refer the reader to Chapters 3 and 4 in Hayashi (2000). For their adaptations to multilevel models and formulas for model specification tests and robust estimation, the reader is referred to Kim and Frees (2007).

11.4.1 The GMM Estimator

Exogenous variables that are useful for estimating coefficients in $\boldsymbol{\beta}$ are said to be *instruments*. When all model variables $\mathbf{X}$ are exogenous, we can estimate $\boldsymbol{\beta}$ exclusively based on $\mathbf{X}$. However, if some $\mathbf{X}$ are endogenous, the econometric IV methods require additional nonmodel instrumental variables for consistent estimation.

On the other hand, Kim and Frees (2007) developed a procedure of building "internal" instruments by utilizing the nested structure of hierarchical data that does not require additional "external" variables. In their approach, internal instruments are functions of variables in the model, and various sets of instruments can be defined in multilevel models.

To define the GMM estimator, we recall the stacked version of a linear mixed effects model $\mathbf{y}_s = \mathbf{X}_s\boldsymbol{\beta} + \boldsymbol{\delta}_s$ in Equation 11.8. The outcome $\mathbf{y}_s$ is independent over s with $\mathrm{E}\boldsymbol{\delta}_s = 0$ and Var $\boldsymbol{\delta}_s = \mathbf{V}_s$. As schools are independent units in this model, the variance-covariance matrix $\mathbf{V}$ is a block diagonal matrix, denoted as $\mathbf{V} = blkdiag$ $(\mathbf{V}_{1,\ldots,}\mathbf{V}_n)$. Let $\mathbf{W}_s$ be a known weight matrix associated with $\boldsymbol{\delta}_s$. The matrix of instruments, denoted as $\mathbf{H}_s$, are exogenous and thus uncorrelated with the random effects $\boldsymbol{\delta}_s$ in the model such that:

$$E\,\mathbf{H}_s'\mathbf{W}_s\boldsymbol{\delta}_s = 0, \quad s = 1,\ldots,n. \tag{11.12}$$

Let $\mathbf{y} = (\mathbf{y}_1',\ldots,\mathbf{y}_n')'$, $\mathbf{X} = (\mathbf{X}_1',\ldots,\mathbf{X}_n')'$, $\mathbf{H} = (\mathbf{H}_1',\ldots,\mathbf{H}_n')'$ and $\boldsymbol{\delta} = (\boldsymbol{\delta}_1',\ldots,\boldsymbol{\delta}_n')'$. With the matrix of weights $\mathbf{W} = blkdiag$ $(\mathbf{W}_{1,\ldots,}\mathbf{W}_n)$, the GMM estimator is defined as:

$$\mathbf{b}_{\mathrm{GMM}} = \left(\mathbf{X}'\mathbf{W}P(\mathbf{H})\mathbf{W}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{W}P(\mathbf{H})\mathbf{W}\mathbf{y}, \tag{11.13}$$

where $P(\mathbf{H}) = \mathbf{H}(\mathbf{H}'\mathbf{H}) - \mathbf{H}'$ is the projection onto the linear space spanned by the columns of $\mathbf{H}$ and "–" denotes a generalized inverse.

For this GMM estimator, we have introduced weights $\mathbf{W}_s$ to allow for a variance structure, which is usually ignored in IV methods. In that case, one may use the identity for the weight matrix. Alternatively, the weight can be the inverse of the square root of the variance–covariance matrix of the disturbance term ($\mathbf{V}_s^{-1/2}$), thus producing a GLS estimate. As another option, in mixed linear effects modeling (see, for example, Diggle, Heagerty, Liang, & Zeger, 2002), it is customary for analysts to use a weight matrix that approximates $\mathbf{V}_s^{-1/2}$ and then use a robust estimate of standard errors to correct for misspecifications:

$$\begin{aligned}\widehat{Var\,\mathbf{b}_{\mathrm{GMM}}} &= \left(\mathbf{X}'\mathbf{W}P(\mathbf{H})\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}P(\mathbf{H})\mathbf{W} \\ &\quad\times\left(blkdiag(\mathbf{e}_1\mathbf{e}_1',\ldots,\mathbf{e}_n\mathbf{e}_n')\right) \\ &\quad\times \mathbf{W}P(\mathbf{H})\mathbf{W}\mathbf{X}'\left(\mathbf{X}'\mathbf{W}P(\mathbf{H})\mathbf{W}\mathbf{X}\right)^{-1},\end{aligned}$$

determined from the residuals $\mathbf{e}_s = \mathbf{y}_s - \mathbf{X}_s\mathbf{b}_{\mathrm{GMM}}$. We used these *empirical robust* standard errors, also known as Huber–White *sandwich* standard errors (Huber, 1967; White, 1982), instead of model-based standard errors in our applications, as the latter are known to be sensitive to omitted variable effects and may provide severely underestimated standard errors and consequently falsely large effects (Kim & Frees, 2006; Maas & Hox, 2004).

11.4.2 Creating the Estimator Continuum Using Internal Instruments

The formula for the GMM estimator in Equation 11.13 indicates that we obtain different $\mathbf{b}_{\mathrm{GMM}}$ with a different set of instruments $\mathbf{H}$. In fact, we can estimate $\mathbf{b}_{\mathrm{FEt}}$, $\mathbf{b}_{\mathrm{FEs}}$, and $\mathbf{b}_{\mathrm{RE}}$ in previous sections using Equation 11.13, and the calculation of FE estimators is much faster than the dummy variable approach mentioned earlier. More importantly, the GMM estimator provides useful alternative options when the RE estimator is biased and the FE estimators are not desirable. Using the FE, RE, and additional GMM estimators, we can create a continuum of estimators from the most robust but least efficient $\mathbf{b}_{\mathrm{FEt}}$ at one end to the most efficient but least robust $\mathbf{b}_{\mathrm{RE}}$ at the other end. The purpose of building this continuum is to find an estimator that is robust against omitted variables and also as efficient as possible.

The internal instruments $\mathbf{H}$ can be constructed without nonmodel variables, consisting of functions of model variables $\mathbf{X}$. In particular, within-group deviations and group means of $\mathbf{X}$ can be used as instruments in multilevel models. Note that different groups can be identified in three- or higher-level models. Previously we considered a school–teacher–pupil model, and schools and teachers can thus be considered as groups having different deviations and means. Also, different amounts of information can be used for different estimators. Specifically, $\mathbf{b}_{\mathrm{FEt}}$ is obtained under the idea that all predictors might be endogenous, and thus corresponding instrument $\mathbf{H}_{\mathrm{FEt}}$ is composed of only within-teacher deviations, which are not affected by omitted variables $\mathbf{u}_s^{(3)}$ and $\mathbf{u}_{st}^{(2)}$. Similarly, the instrument $\mathbf{H}_{\mathrm{FEs}}$ for $\mathbf{b}_{\mathrm{FEs}}$ is composed of within-school deviations.[1]

On the other hand, $\mathbf{b}_{\mathrm{RE}}$ is obtained by assuming all predictors are exogenous and $\mathbf{H}_{\mathrm{RE}}$ is composed of within-group deviations as well as the group means of all predictors in the model. Using more information than $\mathbf{b}_{\mathrm{FEt}}$ and $\mathbf{b}_{\mathrm{FEs}}$, $\mathbf{b}_{\mathrm{RE}}$ is more efficient if all predictors are indeed exogenous. Unlike deviations, however, group means are affected by omitted variables and $\mathbf{b}_{\mathrm{RE}}$ would be biased if some predictors are endogenous. Either teachers or schools as groups provide the same $\mathbf{b}_{\mathrm{RE}}$.

There is a third option for the GMM estimator, taking a middle-of-the-road approach by using both deviations and means of exogenous variables and only deviations of endogenous variables. To implement the GMM estimators, therefore, one needs to separate out potentially endogenous variables in the model, and then construct instruments $\mathbf{H}_{\mathrm{GMM}}$ by merging the within-group deviations of all predictors and the group means of exogenous variables. This approach is reasonable because one would not wish to rely on potentially endogenous variables as much as the other exogenous variables for the estimation of model parameters.

For example, if it is of concern that SES, but not MALE, might be endogenous due to unobserved school variables such as district and neighborhood characteristics, only the within-school deviation $(\mathrm{SES}_{stp} - \overline{\mathrm{SES}}_s)$ may be used as an instrument, whereas both school average $\overline{\mathrm{MALE}}_s$ and within-school deviation $(\mathrm{MALE}_{stp} - \overline{\mathrm{MALE}}_s)$ may be included in $\mathbf{H}_{\mathrm{GMMs}}$ for the estimation of the GMM school estimator $\mathbf{b}_{\mathrm{GMMs}}$. Analogously, teacher deviations and means can be used as instruments for the GMM teacher estimator $\mathbf{b}_{\mathrm{GMMt}}$. The idea of the GMM estimator is rooted in the Hausman–Taylor estimator (Hausman & Taylor, 1981) and IV estimators in econometrics (Hsiao, 2003; Wooldridge, 2002).

To summarize, it may be more than is necessary to regard all predictors as endogenous yet too lenient to assume all are exogenous in many applications. The GMM estimator encompasses the two extreme approaches by treating part of the variables as endogenous. In the school–teacher–pupil model, within-teacher deviations and within-school deviations are used as instruments for $\mathbf{b}_{\mathrm{FEt}}$ and $\mathbf{b}_{\mathrm{FEs}}$, respectively. For the corresponding GMM estimators, denoted as $\mathbf{b}_{\mathrm{GMMt}}$ and $\mathbf{b}_{\mathrm{GMMs}}$, teacher means and school means of

[1] The same idea of using deviations to remove unobserved variable u_i from a linear longitudinal model is shown from Equations 11.3 to 11.5.

exogenous variables are additionally used as instruments. Note that different GMM estimators would be obtained for different choices of endogenous variables. When all predictors are considered as exogenous, the GMM estimator is equivalent to the RE estimator. The five estimators can be arranged in the order of $\mathbf{b}_{\mathrm{FEt}}$, $\mathbf{b}_{\mathrm{GMMt}}$, $\mathbf{b}_{\mathrm{FEs}}$, $\mathbf{b}_{\mathrm{GMMs}}$, and $\mathbf{b}_{\mathrm{RE}}$, with regard to their robustness against omitted variables $\mathbf{u}_s^{(3)}$ and $\mathbf{u}_{st}^{(2)}$.

11.4.3 Comparing Multilevel Model Estimators Using the GMM Tests

After building the estimator continuum, it is of interest to examine if differences among the estimators are statistically significant. The GMM tests for comparing FE, RE, and GMM estimators can be presented as:

$$\chi^2_{\mathrm{GMM}} = (\mathbf{b}_2 - \mathbf{b}_1)' \left(\mathrm{Var}(\mathbf{b}_2 - \mathbf{b}_1)\right)^{-1} (\mathbf{b}_2 - \mathbf{b}_1). \tag{11.14}$$

This test statistic has an asymptotic chi-square distribution with degrees of freedom equal to the rank of Var ($\mathbf{b}_2 - \mathbf{b}_1$). Hausman and colleagues (Hausman, 1978; Hausman & Taylor, 1981) introduced a test statistic of this form for model specification tests with panel data models. Kim and Frees (2006, 2007) extended the Hausman test to more general multilevel models and also incorporated GMM estimators. As a result, the following five types of tests can be conducted using Equation 11.14:

1. Fixed effects estimator versus random effects estimator: for example, $\mathbf{b}_{\mathrm{FEt}}$ versus $\mathbf{b}_{\mathrm{RE}}$
2. Two fixed effects estimators: for example, $\mathbf{b}_{\mathrm{FEt}}$ versus $\mathbf{b}_{\mathrm{FEs}}$
3. GMM estimator versus random effects estimator: $\mathbf{b}_{\mathrm{GMMt}}$ versus $\mathbf{b}_{\mathrm{RE}}$
4. Two GMM estimators: for example, $\mathbf{b}_{\mathrm{GMMt}}$ versus $\mathbf{b}_{\mathrm{GMMs}}$
5. Fixed effects estimator versus GMM estimator: $\mathbf{b}_{\mathrm{FEt}}$ versus $\mathbf{b}_{\mathrm{GMM}}$

To summarize the model specification tests discussed in this chapter, when the model consists of only two levels and no random slopes, the Hausman test in Equation 11.6 is sufficient. For three or more levels, one should consider the multilevel omitted variable tests in Equation 11.11. When the results of the omitted variable tests indicate bias in random effects estimators, one may consider the GMM tests in Equation 11.14. When comparing estimators using the GMM tests, one can also consider each coefficient with a chi-square distribution with one degree of freedom. This is referred to as the *individual coefficient test* or *one-degree-of-freedom test* in Kim and Frees (2007). Individual coefficient tests can be particularly useful in understanding the sources of omitted variable bias.

11.5 AN EXAMPLE WITH NELS:88

The methodology in this chapter can be applied to various forms of multilevel models, including latent growth models for longitudinal data and cross-sectional hierarchical models for organizational data. This section illustrates the implementations of the methods with a three-level model for the National Education Longitudinal Study of 1988 (NELS:88), which is one of the most used large-scale data sets in education.

11.5.1 Selection of Variables and Multiple Imputation

Achievement test scores in the NELS:88 consist of four subjects—mathematics, reading, history, and science—and mathematics is by far the most studied subject among the four. This may be partly due to an understanding that students' mathematics performance is more sensitive to teacher and school effectiveness than other subjects (Shouse & Mussoline, 2002). We also considered the 10th grade mathematics achievement test scores as the outcome variable.

For the selection of predictors, we reviewed the relevant literature extensively to gather variables that are commonly used as predictors in educational production functions. Among many references, we chose most of the variables based on Goldhaber and Brewer (1997), Brewer and Goldhaber (2000), Ehrenberg, Brewer, Gamoran, and Willms (2001), and Rivkin, Hanushek, and Kain (2005) that are hypothesized to be related to student achievement, including teacher and school characteristics as well as student background variables. Wayne and Youngs (2003) surveyed a large number of articles on teacher effectiveness and argued the importance of controlling for prior student achievement. Battistich Solomon, Kim, Watson, and Schaps (1995) and Lee and Smith (1997) also stated that socio-economic status (SES), gender, minority status and some form of prior achievement are necessities in educational production functions. We used the IRT-equated eighth grade mathematics score for controlling for prior achievement.

Table 11.1 shows the predictors at the student, teacher, and school levels and their summary statistics. Note that we compared descriptive statistics based on two different forms of NELS:88; one after listwise deletion ($N = 5278$) by removing all subjects with missing values for the predictors in the model, and the other with imputed observations for the missing values ($N = 7334$) using the multiple imputation procedure implemented in SAS (SAS Institute Inc., 2004; Schafer, 1999). By comparing the percentages of different categories and mean mathematics scores of the variables between the two forms, especially with respect to SES, prior achievement, and minority status, it is apparent that the missing completely at random (MCAR) assumption is far from satisfied in NELS:88 and thus listwise deletion is inappropriate. On the other hand, although the multiple imputation procedure is not assumption free, it requires the considerably weaker assumption of missing at random (MAR) instead of MCAR (Little & Rubin, 2002). Therefore, we used the imputed data set for our analysis in this chapter.

11.5.2 Omitted Variable Tests

We fitted a three-level random intercept model and first obtained the random effects estimator $\mathbf{b}_{\mathrm{RE}}$, assuming no omitted variable effects. This is equivalent to assuming $\mathbf{u}_s^{(3)} = \mathbf{u}_{st}^{(2)} = 0$ in Equation 11.10. As noted earlier, we would not be able to tell if this $\mathbf{b}_{\mathrm{RE}}$ is biased unless we compare it with more robust estimators. Therefore, we also obtained the fixed teacher effects estimator $\mathbf{b}_{\mathrm{FEt}}$, which is robust against the presence of both $\mathbf{u}_s^{(3)}$ and $\mathbf{u}_{st}^{(2)}$. In addition, we obtained the fixed school effects estimator $\mathbf{b}_{\mathrm{FEs}}$ as well, which is robust against the presence of $\mathbf{u}_s^{(3)}$ but not $\mathbf{u}_{st}^{(2)}$.

The three sets of estimates, $\mathbf{b}_{\mathrm{FEt}}$, $\mathbf{b}_{\mathrm{FEs}}$, and $\mathbf{b}_{\mathrm{RE}}$, are listed in Table 11.2. One might question why $\mathbf{b}_{\mathrm{FEt}}$ is not always used if it is robust against both omitted teacher and school effects. The reason is shown in Table 11.2.

TABLE 11.1

Percentages or Means of Predictors and Average Mathematics Scores by Subgroups; Standard Deviations in Parens

	Percentage or Mean		Math Score	
Variable	**Listwise Deletion**	**MI**	**Listwise Deletion**	**MI**
Student				
N	5278	7334		
Current achievement			45.43 (13.74)	44.35 (13.92)
Prior achievement			37.85 (11.96)	36.97 (12.05)
SES	0.07 (0.79)	0.03 (0.80)		
Female	0.51 (0.50)	0.50 (0.50)	45.02 (13.45)	44.12 (13.58)
Male	0.49 (0.50)	0.50 (0.50)	45.84 (14.02)	44.58 (14.24)
Minority	0.24 (0.43)	0.28 (0.45)	40.43 (13.93)	39.11 (13.75)
Caucasian	0.76 (0.43)	0.72 (0.45)	47.01 (13.30)	46.37 (13.44)
Teacher				
N	2151	3016		
Has a math background	0.28 (0.45)	0.27 (0.44)	46.30 (12.46)	45.51 (12.53)
No math background	0.72 (0.45)	0.73 (0.44)	43.41 (13.24)	42.14 (13.30)
Experienced (3 + years)	0.89 (0.31)	0.89 (0.32)	44.67 (13.09)	43.40 (13.24)
Not experienced (< 3 years)	0.11 (0.31)	0.11 (0.32)	40.93 (12.57)	39.98 (12.38)
Female	0.48 (0.50)	0.49 (0.50)	44.20 (13.06)	42.84 (13.18)
Male	0.52 (0.50)	0.51 (0.50)	44.31 (13.11)	43.21 (13.19)
Minority	0.09 (0.28)	0.10 (0.31)	38.80 (12.57)	37.26 (12.30)
Caucasian	0.91 (0.28)	0.90 (0.31)	44.77 (13.01)	43.69 (13.12)
School				
N	626	859		
Urban	0.36 (0.48)	0.37 (0.48)	45.36 (10.96)	43.33 (11.61)
Rural	0.27 (0.44)	0.26 (0.44)	42.53 (8.82)	41.47 (8.92)
Suburban	0.37 (0.48)	0.37 (0.48)	45.58 (9.89)	44.11 (10.04)
School size/100	11.48 (6.70)	11.85 (6.78)		
% Caucasian/10	7.11 (2.77)	6.87 (2.88)		
% Single parent homes/10	2.83 (1.78)	2.94 (1.79)		
Public school	0.82 (0.39)	0.83 (0.38)	43.02 (9.34)	41.48 (9.64)
Private school	0.18 (0.39)	0.17 (0.38)	52.18 (9.99)	51.21 (10.41)

Note: MI: Multiple imputation.

While $\mathbf{b}_{RE}$ provides estimates for all three levels, $\mathbf{b}_{FEt}$ only provides estimates for the level-1 variables and $\mathbf{b}_{FEs}$ provides estimates for the level-1 and level-2 variables. Because of this severe loss of information, one would not choose $\mathbf{b}_{FEt}$ nor $\mathbf{b}_{FEs}$, unless $\mathbf{b}_{RE}$ is biased.

Figure 11.1 provides a flowchart summarizing different paths for testing omitted variable effects and for finding the optimal estimator. Among the three omitted variable tests at the top of Figure 11.1, we started with the omnibus multiple-level test (Test 1)

TABLE 11.2

Omitted Variable Tests; Empirical Standard Errors in Parens

Variable	$\mathbf{b}_{FEt}$	$\mathbf{b}_{FEs}$	$\mathbf{b}_{RE}$	$\mathbf{b}_{FEt}$ vs. $\mathbf{b}_{RE}$	$\mathbf{b}_{FEt}$ vs. $\mathbf{b}_{FEs}$
Intercept			9.71 (0.77)*		
Prior achievement	0.85 (0.01)*	0.94 (0.01)*	0.95 (0.01)*	−0.10 (0.01)*	−0.09 (0.01)*
SES	0.71 (0.18)*	0.94 (0.15)*	1.11 (0.13)*	−0.40 (0.13)*	−0.23 (0.12)*
Female	0.14 (0.21)	0.17 (0.18)	0.17 (0.17)	−0.04 (0.14)	−0.03 (0.13)
Minority	−0.55 (0.31)	−0.64 (0.25)*	−0.76 (0.23)*	0.21 (0.20)	0.10 (0.18)
Math background		0.71 (0.25)*	0.53 (0.21)*		
Experienced		0.56 (0.36)	0.45 (0.30)		
Female		0.42 (0.25)	0.39 (0.20)*		
Minority		−0.36 (0.46)	−0.42 (0.38)		
Urban			−0.43 (0.28)		
Rural			−0.37 (0.28)		
School size/100			−0.02 (0.02)		
% Caucasian/10			−0.01 (0.05)		
% Single parent homes/10			−0.10 (0.07)		
Public school			−0.93 (0.37)*		

*$p < 0.05$.
Omitted variable test 1: H_0: No omitted teacher and school effects exist. $\mathbf{b}_{FEt}$ vs. $\mathbf{b}_{RE}$.
$\chi^2 = 143.34$, $df = 4$, $p < 0.01$.
Omitted variable test 2: H_0: No omitted teacher effects exist. $\mathbf{b}_{FEt}$ vs. $\mathbf{b}_{FEs}$.
$\chi^2 = 130.64$, $df = 4$, $p < 0.01$.

for examining omitted school and teacher effects simultaneously. It is generally suggested to start with Test 1 if omitted variable effects are of concern at multiple levels. The other tests can be subsequently conducted as shown in Figure 11.1. One may also start with the intermediate-level test (Test 2), if the analyst wishes to test omitted variable effects at each level separately. However, caution should be made before starting with the highest-level test (Test 3), as it is not valid if there exist omitted variables at lower levels. See Section 11.3 for further discussion on the properties of the three OV tests.

The results of the OV tests are summarized at the bottom of Table 11.2. On the basis of the multiple-level test, it is clear that $\mathbf{b}_{RE}$ is biased. Even with the empirical standard errors in Section 11.1 (as opposed to model-based errors that tend to underestimate variability and may provide falsely large test statistics), the chi-square test statistic is very large (143.34, $df = 4$, $p < 0.01$). Using the individual coefficient tests, we found statistically significant differences for the effects of prior mathematics achievement (0.85 vs. 0.95) and SES (0.71 vs. 1.11). This means the effects of the two variables would be upward biased, if we ignore omitted variable effects and use $\mathbf{b}_{RE}$.

Next, we conducted the intermediate-level test and the results indicate that there exist significant omitted teacher effects ($\chi^2 = 130.64$, $df = 4$, $p < 0.01$) and that $\mathbf{b}_{FEs}$ is also biased. As we found that $\mathbf{b}_{FEs}$ is biased, comparing the biased estimator to another biased estimator

$\mathbf{b}_{RE}$ is not meaningful. Therefore, we did not conduct the highest-level test. This example shows the very reason why the econometric treatment for two-level models discussed in Section 11.2 would fall short in multilevel models. If we had applied the Hausman test to examine omitted school effects at the third level (i.e., OV Test 3), the test would have compared two biased estimators.

Although we have one estimator that is robust to the omitted teacher and school effects, we cannot estimate the effects of teacher-level and school-level variables using $\mathbf{b}_{FEt}$. Sweeping out higher-level effects is a critical limitation of fixed effects approaches and might be one of the main reasons why efficient random effects estimators have been used routinely despite the danger of omitted variable bias. To avoid this problem, some studies selected out students who switched schools, which allow for estimating the effects of teacher and school variables. However, this approach has several serious problems. Most importantly, these "movers" or "switchers" are usually a small proportion of the whole population and often have different characteristics than the rest. Consequently, not only can the estimation be unstable due to a small sample size, but also the findings based on the movers' group may not be generalizable to the majority who did not switch schools. Therefore, instead of relying on biased $\mathbf{b}_{RE}$ or limited information $\mathbf{b}_{FEt}$, we further our analysis to find the GMM estimator.

11.5.3 Searching for the Optimal Estimator

As explained earlier, the RE and FE estimators treat all and none of predictors as exogenous, respectively. On the other hand, the GMM estimator provides a middle ground between the two extreme "all or nothing" approaches and allows for some of predictors to be endogenous and treats the others as exogenous. By making this distinction, the analyst can use both the within- and between-group information of exogenous variables, while using only the within-group information from endogenous variables as instruments (see Section 11.2).

In this chapter, we demonstrate the internal instruments approach that utilizes variables already in the model. If available, external variables can be easily added to the set of instruments. However, it is important that one can obtain the unbiased GMM estimator without relying on additional variables that may not be available. It is also important to identify endogenous variables properly. We expect that the analyst would have some understanding or theory about endogeneity issues in relation to the nature of the topic (e.g., confounding factors) and the data (e.g., data collection process), based on experience and knowledge in the field. The choices of instruments should be guided by this knowledge.

It should be clarified that the purpose of our example is to demonstrate the methodologies rather than to make substantive inferences. Nonetheless, we made efforts to gather proper information and our choice of predictors and endogenous variables are based on the literature in the field. Among the predictors summarized in Table 11.1, prior achievement, SES, and school-level percentage of minority students were suspected to be endogenous. Prior achievement is a concern in regard to factors that would affect both prior and current achievement but are not available in the data set. It is natural to view any value-added type predictor as endogenous because of this inevitable relationship (Alexander, Pallas, & Cook, 1981).

SES would also likely be correlated with numerous factors that influence student achievement. Finally, there is concern that school-level percentage of minority students might be related to omitted school, district, and neighborhood information.

Thus, we hold the three predictors as potential endogenous variables, which means only within-teacher deviations of these variables are used as instruments, whereas both within-teacher deviations and teacher-level means of the rest of the 14 variables are used as instruments **H** in the calculation of $\mathbf{b}_{\text{GMMt}}$ in Equation 11.13. Similarly, within-school deviations of the three endogenous variables are used and both within-school deviations and school-level means of the exogenous variables are used as instruments for $\mathbf{b}_{\text{GMMs}}$. Although we obtained $\mathbf{b}_{\text{GMMs}}$ to complete the estimator continuum, as we already found that $\mathbf{b}_{\text{FEs}}$ is biased in the OV Test 2 above, the corresponding GMM estimator does not carry much value in the current analysis. The results of the five estimators are summarized in Table 11.3. They are ordered from the most robust to the least robust as: $\mathbf{b}_{\text{FEt}}$, $\mathbf{b}_{\text{GMMt}}$, $\mathbf{b}_{\text{FEs}}$, $\mathbf{b}_{\text{GMMs}}$, and $\mathbf{b}_{\text{GLS}}$.

An important question in Table 11.3 is whether $\mathbf{b}_{\text{GMMt}}$ is as robust as $\mathbf{b}_{\text{FEt}}$. If so, we can obtain unbiased estimates for the effects of variables at all levels from $\mathbf{b}_{\text{GMMt}}$. This constitutes the GMM Test 5 in Section 11.3, a test between an FE estimator and a GMM estimator. We found that $\mathbf{b}_{\text{FEt}}$ and $\mathbf{b}_{\text{GMMt}}$ provide almost identical estimates and the test statistic (χ^2) was as small as 0.03. Therefore, we concluded that $\mathbf{b}_{\text{GMMt}}$ is the optimal estimator in our three-level model for the mathematics achievement scores in NELS:88.

While making inferences based on $\mathbf{b}_{\text{GMMt}}$, we also compared $\mathbf{b}_{\text{GMMt}}$ to $\mathbf{b}_{\text{RE}}$ for each coefficient. This comparison may help understand sources of omitted variable bias in $\mathbf{b}_{\text{RE}}$, and the information can be utilized in further research. The detailed results of this comparison are shown in Table 11.4. The test between $\mathbf{b}_{\text{GMMt}}$ and $\mathbf{b}_{\text{RE}}$ results in a large χ^2 value of 150.55 ($p < 0.01$) with 15 degrees of freedom. The individual coefficient tests showed significant differences for many variables including prior achievement, family SES, urbanity, school size, school percentage of Caucasian, and public status, suggesting the omission of some important information in relation to these variables. This is consistent with our literature review and concern over the importance of accounting for individual and family attributes, community dynamics, and school climates in the study of school and teacher effectiveness (e.g., Alexander et al., 1981).

Interestingly, there was a clear pattern in the direction of bias. While $\mathbf{b}_{\text{RE}}$ overestimates the student level variables of prior achievement and family SES, it substantially underestimates several school effects, especially concerning characteristics of schools such as urban status, public status, school size, and ethnicity composition. Based on the robust estimator $\mathbf{b}_{\text{GMMt}}$, the negative effect of students' ethnicity (being minority) became nonsignificant, while several school-level variables, including school size and the percentage of Caucasian, became significant.

We conclude our data analysis example with suggesting several general guidelines for the analysis of other data sets, which can be accompanied by the flowchart in Figure 11.1. First, when $\mathbf{b}_{\text{RE}}$ is biased, one may end the analysis with $\mathbf{b}_{\text{FEt}}$ or $\mathbf{b}_{\text{FEs}}$, especially if it is difficult to identify endogenous variables or it is suspected that the majority of predictors are endogenous. In many applications, however, the GMM estimator

TABLE 11.3

Five Estimators From Most Robust to Most Efficient; Empirical Standard Errors in Parens

Variable	$\mathbf{b}_{FEt}$	$\mathbf{b}_{GMMt}$	$\mathbf{b}_{FEs}$	$\mathbf{b}_{GMMs}$	$\mathbf{b}_{RE}$
Intercept		9.41 (1.57)*		8.97 (0.99)*	9.71 (0.77)*
Prior achievement	0.85 (0.01)*	0.85 (0.01)*	0.94 (0.01)*	0.94 (0.01)*	0.95 (0.01)*
SES	0.71 (0.18)*	0.71 (0.17)*	0.94 (0.15)*	0.94 (0.15)*	1.11 (0.13)*
Female	0.14 (0.21)	0.12 (0.17)	0.17 (0.18)	0.17 (0.17)	0.17 (0.17)
Minority	−0.55 (0.31)	−0.55 (0.31)	−0.64 (0.25)*	−0.68 (0.25)*	−0.76 (0.23)*
Math background		0.58 (0.23)*	0.71 (0.25)*	0.52 (0.21)*	0.53 (0.21)*
Experienced		0.59 (0.31)	0.56 (0.36)	0.47 (0.30)	0.45 (0.30)
Female		0.39 (0.21)	0.42 (0.25)	0.40 (0.20)*	0.39 (0.20)*
Minority		0.16 (0.48)	−0.36 (0.46)	−0.27 (0.40)	−0.42 (0.38)
Urban		0.18 (0.38)		−0.28 (0.30)	−0.43 (0.28)
Rural		−0.54 (0.32)		−0.39 (0.28)	−0.37 (0.28)
School size/100		0.08 (0.03)*		0.04 (0.02)	0.02 (0.02)
% Caucasian/10		0.52 (0.15)*		0.12 (0.08)	0.01 (0.05)
% Single parent homes/10		−0.05 (0.10)		−0.08 (0.07)	−0.10 (0.07)
Public school		−1.94 (0.42)*		−1.07 (0.38)*	−0.93 (0.37)*

*$p < 0.05$.

can be as robust as the corresponding FE estimator and would be preferred. Second, one may choose $\mathbf{b}_{FEs}$ over $\mathbf{b}_{FEt}$, as long as the difference is not significant. Similarly, $\mathbf{b}_{GMMs}$ would be preferred to $\mathbf{b}_{GMMt}$, if both are unbiased. Third, consider the comparison between $\mathbf{b}_{FEs}$ and $\mathbf{b}_{GMMs}$ in the right-most panel. If the difference is significant (i.e., $\mathbf{b}_{GMMs}$ is biased), one may further consider $\mathbf{b}_{GMMt}$ instead of $\mathbf{b}_{FEs}$. Note that $\mathbf{b}_{FEs}$ and $\mathbf{b}_{GMMt}$ are not directly comparable so $\mathbf{b}_{GMMt}$ should be compared to $\mathbf{b}_{FEt}$. If the test between $\mathbf{b}_{FEt}$ and $\mathbf{b}_{GMMt}$ reveals that $\mathbf{b}_{GMMt}$ is biased, one would go back to $\mathbf{b}_{FEs}$ in the previous step. In sum, the flowchart starts with one of the three OV tests and ends with one of the five estimators for a three-level model. There would be a larger number of entities and paths for higher-level models. In our analysis of NELS:88, the starting point was the OV Test 1 and the ending point was $\mathbf{b}_{GMMt}$ in the middle panel.

11.6 SUMMARY AND DISCUSSION

Although few would argue the danger of omitted variable bias, the harmful consequences in data analysis are often overlooked. This is partly because of the lack of statistical methods for handling omitted variables in multilevel models until recently. As in many observational studies the analyst does not have the ability to collect all the "right" variables, it is of great interest to utilize statistical techniques to handle omitted variables as much as possible and ideally obtain unbiased solutions.

This chapter provides an introduction to recent statistical methodology for model specification tests and robust estimation techniques in multilevel models including three types of omitted variable tests (Section 11.3) and GMM estimators (Section 11.4).

TABLE 11.4

Comparing the Teacher-Level GMM Estimator and the Random Effects Estimator

Variable	$\mathbf{b}_{GMMt}$	$\mathbf{b}_{RE}$	Difference	Std Err	Individual Coeff. Test (χ_1^2)
Prior achievement	0.85	0.95	−0.10	0.01	108.70, $p < 0.01$
SES	0.71	1.11	−0.40	0.13	9.85, $p < 0.01$
Female	0.12	0.17	−0.06	0.08	0.42, $p = 0.52$
Minority	−0.55	−0.76	0.21	0.20	1.11, $p = 0.29$
Math background	0.58	0.53	0.06	0.20	0.20, $p = 0.65$
Experienced	0.59	0.45	0.14	0.16	0.73, $p = 0.39$
Female	0.39	0.39	0.00	0.10	0.00, $p = 0.99$
Minority	0.16	−0.42	0.57	0.30	3.55, $p = 0.06$
Urban	0.18	−0.43	0.60	0.27	5.04, $p = 0.02$
Rural	−0.54	−0.37	−0.17	0.16	1.15, $p = 0.28$
School size/100	0.08	0.02	0.06	0.02	11.69, $p < 0.01$
% Caucasian/10	0.52	0.01	0.51	0.13	14.20, $p < 0.01$
% Single parent homes/10	−0.05	−0.10	0.05	0.08	0.44, $p = 0.51$
Public school	−1.94	−0.93	−1.02	0.21	23.72, $p < 0.01$

GMM Test: $\mathbf{b}_{GMMt}$ vs. $\mathbf{b}_{RE}$
$\chi^2 = 150.55$, $df = 15$, $p < 0.01$

It is shown that the versatile GMM technique provides an overarching framework encompassing the well-known random and fixed effects estimators and also offers additional and often more desirable options between the two extremes.

In the three-level model analysis for NELS:88, $\mathbf{b}_{RE}$ turned out to be severely biased. However, we found $\mathbf{b}_{GMMt}$, which is unbiased without losing higher-level information like $\mathbf{b}_{FEt}$. Despite its advantageous properties, an outstanding shortcoming of GMM estimation is that its implementation is cumbersome, as formulas for the GMM estimator in Equation 11.13 and the GMM tests in Equation 11.14 are not utilized in statistical programs. All required formulas are given in Kim and Frees (2007). The SAS IML code written by Frees, Kim, and Swoboda is available by request to the authors of this chapter.

It is well known that multilevel models can be written as linear mixed-effects models. However, this chapter demonstrates that it is critical to retain the multiple-level representation when inspecting omitted variables at different levels. Also, the GMM estimation technique exploits the hierarchical nature of multilevel data and can create internal instruments, so that the researcher is not forced to look for additional variables that were not involved in the original model formulation. For nonhierarchical data (without replications within clusters), one cannot obtain unbiased GMM estimators without external instruments in the presence of endogenous variables.

Additionally, for those who are familiar with IV approaches in econometrics, the GMM methodology in this chapter extends that related work in several important ways. First, the GMM estimator generalizes the original work of Hausman and Taylor (1981) for panel data models to more complex multilevel frameworks. Second, the GMM tests provide a general procedure for directly comparing various types of estimators beyond the FE and RE estimators. Third, empirical standard errors are adapted as opposed to traditional model-based standard errors that are known to underestimate variability when the models are not correctly specified. Finally, the GMM estimator extends the IV estimator by incorporating weights to accommodate the variance structure of a multilevel model and can handle more complex covariance structures in hierarchical data.

As a final note, we recall that Kim and Frees (2006) linked the omitted variable problem to a larger issue of unobserved heterogeneity in the population. Unobserved heterogeneity is a recurrent issue across many disciplines, including econometrics, psychometrics, biostatistics, and sociology. However, the commonality across these literatures has been overlooked, and problems related to unobserved heterogeneity have been acknowledged under various names such as latent classes or finite mixtures, omitted variables, correlated effects, unobserved covariates, measurement error, and confounding variables. In these applications, different assumptions are made about the nature of the unobserved variables (e.g., mutual exclusiveness, independent error terms, time-constant or time-varying variables, parametric or nonparametric mixing distributions) and also different implications of unobserved heterogeneity are emphasized in different disciplines (e.g., impact on causal inference, bias in regression coefficients, collapsibility, etc.). Among the extensive literature dealing with these issues, we refer to Heckman and Singer (1982), Chamberlain (1985), Yamaguchi (1986), Palta and Yao (1991), Vermunt (1997), Frank (2000), McLachlan and Peel (2000), Halaby (2004), and Frees (2004) for further readings.

ACKNOWLEDGMENT

We are thankful to Dan Bolt and David Kaplan for their useful comments.

REFERENCES

Alexander, K. L., Pallas, A. M., & Cook, M. A. (1981). Measure for measure: On the use of endogenous ability data in school-process research. *American Sociological Review, 46,* 619–631.

Battistich, V., Solomon, D., Kim, D.-I., Watson, M., & Schaps, E. (1995). Schools as communities, poverty levels of student populations and students' attitudes, motives and performance: A multilevel analysis. *American Educational Research Journal, 32,* 627–658.

Brewer, D. J., & Goldhaber, D. D. (2000). Improving longitudinal data on student achievement: Some lessons from recent research using NELS:88. In D. Grissmer & M. Ross (Eds.), *Analytic issues in the assessment of student achievement.* Washington DC: NCES.

Chamberlain, G. (1985). Heterogeneity, omitted variable bias, duration dependence. In J. Heckman & B. Singer (Eds.), *Longitudinal analysis of labor market data.* Cambridge, NY: Cambridge University Press.

Diggle, P. J., Heagarty, P., Liang, K.-Y., & Zeger, S. L. (2002). *Analysis of longitudinal data* (2nd ed.). London: Oxford University Press.

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Willms, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest, 2,* 1–30.

Frank, K. A. (2000). Impact of a confounding variable on a regression coefficient. *Sociological Methods & Research, 29,* 147–194.

Frees, E. W. (2004). *Longitudinal and panel data: Analysis and applications for the social sciences.* Cambridge: Cambridge University Press.

Frees, E. W., & Kim, J.-S. (2008). Panel studies. In T. Rudas (Ed.), *Handbook of probability: Theory and applications* (pp. 205–224). Thousand Oaks, CA: Sage.

Goldhaber, D. D., & Brewer, D. J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *The Journal of Human Resources, 32,* 505–523.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London, UK: Oxford University Press.

Halaby, C. H. (2004). Panel models in sociological research: Theory into practice. *Annual Review of Sociology, 30,* 507–540.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica, 50,* 1029–1054.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica, 46,* 1251–1272.

Hausman, J. A., & Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica, 49,* 1377–1398.

Hayashi, F. (2000). *Econometrics.* Princeton, NJ: Princeton University Press.

Heckman, J. J., & Singer, B. (1982). Population heterogeneity in demographic models. In K. Land & A. Rogers (Eds.), *Multidimensional mathematical demography.* New York, NY: Academic Press.

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Hsiao, C. (2003). *Analysis of panel data* 2nd ed. Cambridge, UK: Cambridge University Press.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability.* Berkeley, CA: University of California Press.

Kim, J.-S. (2009). Multilevel analysis: An overview and some contemporary issues. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *Handbook of quantitative methods in psychology* (pp. 337–361). Thousand Oaks, CA: Sage.

Kim, J.-S., & Frees, E. W. (2006). Omitted variables in multilevel models. *Psychometrika, 71,* 659–690.

Kim, J.-S., & Frees, E. W. (2007). Multilevel modeling with correlated effects. *Psychometrika, 72,* 505–533.

Lee, V. E., & Smith, J. B. (1997). High school size: Which works best and for whom? *Educational Evaluation and Policy Analysis, 19,* 205–227.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: John Wiley.

Maas, C. J. M., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica, 58,* 127–137.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models.* New York, NY: Wiley.

Palta, M., & Yao, T.-J. (1991). Analysis of longitudinal data with unmeasured confounders. *Biometrics, 47,* 1355–1369.

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

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73,* 417–458.

SAS Institute Inc. (2004). *SAS/STAT 9.1 User's Guide.* Cary, NC: SAS Institute Inc.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8,* 3–15.

Shouse, R. C., & Mussoline, L. J. (2002). High risk, low return: The achievement effects of restructuring in disadvantaged schools. *Social Psychology of Education, 3,* 245–259.

Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 24,* 323–355.

Snijders, T. A., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London, UK: Sage.

Vermunt, J. K. (1997). *Log-linear models for event histories.* Thousand Oaks, CA: Sage.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73,* 89–122.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica, 50,* 1–25.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data.* Cambridge, MA: MIT Press.

Yamaguchi, K. (1986). Alternative approaches to unobserved heterogeneity in the analysis of repeatable events. In B. Tuma (Ed.), *Sociological methodology.* Washington, DC: American Sociological Association.