Omitted Variables in Multilevel Models

Jee-Seon Kim

Edward W. Frees

University of Wisconsin—Madison

2006

Psychometrika, in press

This research was supported by National Academy of Education/Spencer Foundation and the National Science Foundation, Grant Number SES-0436274. Correspondence concerning this article should be addressed to Jee-Seon Kim, Department of Educational Psychology, University of Wisconsin, 1025 West Johnson Street, Madison, WI 53706. Electronic mail may be sent to jeeseonkim@wisc.edu.
Omitted Variables in Multilevel Models

Abstract

Statistical methodology for handling omitted variables is presented in a multilevel modeling framework. In many non-experimental studies, the analyst may not have access to all requisite variables, and this omission may lead to biased estimates of model parameters. By exploiting the hierarchical nature of multilevel data, a battery of statistical tools are developed to test various forms of model misspecification as well as to obtain estimators that are robust to the presence of omitted variables. The methodology allows for tests of omitted effects at single and multiple levels. The paper also introduces intermediate-level tests; these are tests for omitted effects at a single level, regardless of the presence of omitted effects at a higher level. A simulation study shows, not surprisingly, that the omission of variables yields bias in both regression coefficients and variance components; it also suggests that omitted effects at lower levels may cause more severe bias than at higher levels. Important factors resulting in bias were found to be the level of an omitted variable, its effect size, and sample size. A real data study illustrates that an omitted variable at one level may yield biased estimators at any level and, in this study, one cannot obtain reliable estimates for school-level variables when omitted child effects exist. However, robust estimators may provide unbiased estimates for effects of interest even when the efficient estimators fail, and the one-degree-of-freedom test helps one to understand where the problem is located. It is argued that multilevel data typically contain rich information to deal with omitted variables, offering yet another appealing reason for the use of multilevel models in the social sciences.

Key words: hierarchical linear models, omitted effects, model specification test, fixed-effects estimators, random-effects estimators, Hausman test, unobserved heterogeneity
1. Introduction

Virtually any statistical modeling approach faces specification problems, and omission of relevant variables is a common difficulty. In ordinary regression models, the consistency of standard least squares estimators depends on the assumption that the explanatory variables are uncorrelated with the disturbance term. This assumption is prone to be violated, especially when important explanatory variables are excluded from the model. Often such omissions are unavoidable due to the inability to collect necessary variables for the model. For example, in a study of the effects of school characteristics on students’ achievement, variables related to teacher preparation and qualifications are known to highly influence students but are often infeasible to collect, especially in large scale datasets. Unfortunately, these variables are likely to be correlated with other explanatory variables in the model, such as per student spending and class size, which are often of interest to educational researchers. A consequence is that not only are we unable to estimate the effects of important variables like teachers’ preparation, but also our estimators for other effects in the model may be biased and thus misleading. This problem is often called an omitted variable bias (see, for example, Chamberlain, 1978).

Although researchers are aware of the danger of omitted variable bias, it is naturally impractical to collect all requisite data in the social and behavioral sciences, as many studies are strictly observational. In cross-sectional or single-level data, the researcher must make strong additional assumptions and use the appropriate (complex) technique to overcome this bias. Fortunately, in the multilevel case, the data contain rich information that provides opportunities to test the severity of bias due to omitted variables as well as to employ alternative estimators that minimize its consequences. This ability to examine omitted variable bias provides yet another compelling reason for using multilevel models. It is the purpose of the current study to demonstrate how to capitalize on these features of hierarchical data.

1.1 Omitted Variable Bias

Despite immense research efforts since the classic Coleman report (Coleman et al., 1966), results concerning the effects of variables such as class size, per-pupil expenditures, teacher qualifications, and administrative factors on educational outcomes remain mixed (Murnane & Phillips, 1981; Hedges, Laine, Greenwald, 1994; Phillips, 1997; Ehrenberg, Brewer, Gamoran, & Willms, 2001; Hanushek, 2003). These mixed findings are likely due to inherent difficulties in isolating the effects of variables involved in complex environments such as schools, but also to factors outside the school (e.g., family background variables, neighborhood variables). Naturally, these problems cannot be totally controlled in any multilevel study.

However, a fundamental potential problem that can be alleviated is that of omitted variable bias. A number of researchers have suggested that the inconsistent findings in school effectiveness research may be due in part to the inappropriateness of the models utilized in the statistical analysis (Goldhaber & Brewer, 1997; Ludwig & Bassi, 1999; Bonesrønning, 2004; Marsh, 2004). As a simple example of omitted variable
bias, consider a setting in which information about a school's socioeconomic status (i.e., the average SES of students within a school) is not collected, but is correlated with the other variables considered in the model (e.g., class size, teacher experience, and the percentage of minority students). In this case, the effects of studied predictors may be over- or under-estimated as a consequence of the omission of school socioeconomic status. In general, when a variable is correlated with predictors included in the model but is excluded from the analysis, the effects of the predictors in the model will not be estimated properly. Conceptually, this makes good sense, as these correlated predictors function as partial surrogates for school mean SES.

The danger of omitted variables has been a recurrent issue in the social sciences. Boardman and Murnane (1979) underscored the potential bias and inconsistency of the ordinary least square estimators, and promoted a panel data approach and estimating equations to overcome the limitations of cross-sectional data and single equation models. Ehrenberg and colleagues incorporated instrumental variable approaches for the analysis of the High School and Beyond (Ehrenberg & Brewer, 1994) and National Education Longitudinal Study of 1988 (Ehrenberg, Goldhaber, & Brewer, 1995). Ehrenberg and Brewer (1995) also reanalyzed the Coleman data using economic models to account for the probable endogeneity of teacher characteristics. Several other studies have considered a variety of procedures to address problems related to omitted variables in education. Rice, Jones, and Goldstein (1998) presented a conditioned iterative generalised least squares estimator to account for correlated effects between random effects and fixed predictors, Dee (1998) used two-stage least squares estimates to determine the quality of public schools, and Rivkin, Hanushek and Kain (2005) used semiparametric lower bound estimates of the variance in teacher quality to disentangle the impact of schools and teachers in influencing academic achievement with the problems of omitted variables.

Despite this attention, the fixed effects estimator approach (e.g., Hausman, 1978) has seldom been used for the analysis of educational multilevel data. This is because fixed effects estimators, although robust, wipe out the effects of many variables in the model. For example, fixed child effects sweep out all child, teacher, and school effects even though the effects of school and teacher characteristics and student background variables are often of primary interest. Therefore, efficient random effects estimators have been used routinely for multilevel models in education, despite potential omitted variable bias. Later in the real data example, we demonstrate how estimators may vary due to an omitted variable.

To illustrate the problem of omitted variables mathematically, consider a “true” model:

\[ y = X \beta + U \gamma + \varepsilon, \]

(1.1)

where \( y \) is the outcome variable, \( X \) are observed and \( U \) are unobserved predictors that affect the outcome, and \( \varepsilon \) is an error term. Since the \( U \) are unobserved—hence omitted in the analysis—the “fitted” model is

\[ y = X \beta + \bar{\varepsilon}, \]
where \( \hat{\varepsilon} = U\gamma + \varepsilon \). The expected value of the least squares estimates for the regression coefficients associated with \( X \) can be shown to be \( \beta + (X'X)^{-1}X'U\gamma \). Unless either \( X'U = 0 \) or \( \gamma = 0 \), the least squares estimator of \( \beta \) is biased and inconsistent.

Thus, the omission of relevant predictor variables causes bias because it induces a correlation between the disturbance term and the explanatory variables. It is well known that correlation between the disturbance term and the explanatory variables may also arise due to (1) measurement errors in the explanatory variables, (2) self-selection (an exogenous choice influences both the dependent and explanatory variables) and (3) simultaneity (the dependent and explanatory variables are jointly determined). Each of these settings can also be motivated as an omitted variable problem. See, for example, a survey paper by Ebbes, et al. (2004) or a standard econometric reference such as Wooldridge (2002).

There are many ways in which omitted variables may enter a problem that are far more complex than the simple linear mechanism described in equation (1.1). To employ procedures that provide some protection against omitted variable bias, analysts must have a sense of the source of potentially important omitted variables.

Although issues related to omitted variables have been systematically studied for (two-level) panel data (largely in the econometrics literature, for example, Hausman, 1978; Hsiao, 2003), the implications of omitted variables for general multilevel models (Snijders & Bosker, 1999; Raudenbush & Bryk, 2002; Goldstein, 2003) have yet to be explored. It might be expected that the consequences of omitted variables will be more complex and dangerous in the multilevel case, as the effects of omitted variables at one level can pervade all levels of the model. Given the growing popularity of multilevel models, the need for a better understanding of these issues seems obvious. In the next section, we provide an introduction to the Hausman test as used in panel data models. In the subsequent sections, we show how this test can be modified and extended to other contextual and longitudinal multilevel designs that are common in the social and behavioral sciences.

1.2. Omitted Variables in Panel Data Models

In econometrics, a panel data model for longitudinal data and can be written as:

\[
y_{it} = \alpha_i + x_{it}' \beta + u_{it} + \varepsilon_{it},
\]

where individual \( i = 1, \ldots, n \) is observed over time points \( t = 1, \ldots, T_i \). Equation (1.2) includes the outcome variable \( y_{it} \) (test score, for example), disturbance term \( \varepsilon_{it} \), explanatory variables \( x_{it} \), and the coefficient vector \( \beta \). The model also contains a latent intercept variable \( \alpha_i \) that is constant over time. This latent variable induces a correlation among individual test scores over time and serves as a proxy for unobserved time-constant characteristics that are uncorrelated with the explanatory variables. Without the omitted variable \( u_{it} \), the model in equation (1.2) is a two-level random intercept model.
Unlike $\alpha_i$, $u_i$ is assumed to be correlated with one or more of the explanatory variables in $x_i$. Thus, this variable creates a bias in the usual least squares estimate of $\beta$. To mitigate the effects of $u_i$, one can apply a fixed effects transformation, sweeping out the time-constant omitted effects. To see the impact of this strategy, take averages over time in equation (1.2) to get

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

Subtracting this from equation (1.2) yields the transformed model equation

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)' \beta + \varepsilon_{it} - \bar{\varepsilon}_i. \quad (1.3)$$

Then the ordinary least squares (OLS) estimator of $\beta$ is

$$b = \left( \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \right)^{-1} \left( \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) \right); \quad (1.4)$$

this estimator is robust to the unobserved variable effects. Note that this estimator assumes that all explanatory variables vary by time; if the $j$th explanatory variable of $x_{it}$ is constant over time, then the $j$th row of $\{x_{it} - \bar{x}_i\}$ is identically zero so that $\beta$ is not estimable.

To determine the importance of omitted variables, we again consider the model in equation (1.2) but now assume that there are no omitted variables so that $u_i = 0$. We retain, however, the random effects $\alpha_i$. Then, estimating $\beta$ via the usual generalized least squares (GLS) routines, we denote the resulting estimator as $b_{\text{REE}}$. The test statistic

$$\chi^2_H = (b - b_{\text{REE}})' (\text{Var} b - \text{Var} b_{\text{REE}})^{-1} (b - b_{\text{REE}})$$

measures the distance between vectors $b$ and $b_{\text{REE}}$. Hausman (1978) showed that $\chi^2_H$ has a chi-square distribution under the null hypothesis of no omitted variables, with degrees of freedom equal to the number of parameters in $\beta$.

This simple setting underscores the major strengths and limitations of the approach advocated in this paper. One important strength is that there are relatively few assumptions needed; we will advocate that these procedures are robust to different types of omitted variables and are easy to understand and interpret by analysts. Throughout, we will see that the robust procedures are essentially determined by “sweeping out” unobserved variables as in Equation (1.3) and then using standard inference procedures. In contrast, an alternative procedure, 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 by the analyst.

Another important advantage is that these procedures are easy to implement. Generally, one can implement the calculations using standard statistical packages with 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.

Equation (1.4) also underscores an important limitation – certain regression coefficients are no longer estimable when relevant omitted variables are swept out. This may be a critical feature of an analysis of a data set – to overcome this drawback additional assumptions may be required. We discuss two modifications that overcome this limitation in Section 4 and comment on this aspect further in Section 7.

The difference between $b_{FE}$ and $b_{REE}$ may be due to omitted variables which can be detected using the Hausman test. Under what types of designs are we likely to observe important differences? In the panel data context, Maddala (1971) observed that one could express

$$b_{REE} = (I - \Delta) b_{FE} + \Delta b_B,$$

where $b_B$ is the between-groups estimator and $\Delta$ is a measure of the relative precision of the estimators. In a balanced design ($T_i = T$), one computes the between-groups estimator as

$$b_B = \left( \sum_{i=1}^{n} (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y}) \right) \left( \sum_{i=1}^{n} (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y}) \right)^{-1}.$$

From this formulation and equation (1.4), it is easy to see that a level-2 omitted variable does not induce a bias in $b_{FE}$ but will affect $b_B$. Further, as the number of replicates ($T$) or the variability at the lower level increases, the difference between $b_{REE}$ and $b_{FE}$ becomes less prominent. Conversely, for a given impact of a level-2 omitted variable, one can anticipate more prominent differences between $b_{REE}$ and $b_{FE}$ when the number of lower level observations per upper level unit is small or the variability at the upper level is small (see, for example, Blundell and Windmeijer, 1997). As we will see, this pattern also holds in higher order designs.

The other important limitation of our proposed procedures is that they do not address so-called “level 1 dependencies.” In the context of the equation (1.2) special case, we assume that the disturbance term $\epsilon_{it}$ is uncorrelated with the explanatory variables. That is, all important omitted variables do not vary over $t$ and may be subsumed with the $u_i$ term. If this assumption is not approximately correct, then other techniques such as the aforementioned instrumental variable or simultaneous equation estimation must be used. See Ebbes et al. (2004) for further discussion of this point in the context of two level panel data models.

Kim and Frees (2005) describes a general two-level model, also known as a linear mixed-effects model, that employs the same strategy. For this general model, we include slopes that vary by subject as well as more general error structures. It is well known that multilevel models may be written as a special case of the linear mixed-effects model; this has been shown useful for parameter estimation purposes by, for example, Singer (1998).
The remainder of the paper is structured as follow: We begin in Sections 2 and 3 by showing how the omitted variable strategy works for a three-level model; Section 2 describes model specification and estimation while Section 3 addresses statistical testing. The issues discussed include the level of the model at which the omitted variable occurs as well as the efficiency of the robust estimation procedure. In the traditional econometrics formulation with only two levels, there is no issue of higher order omitted effects. However, in more general cases, it turns out that the two-level robust estimator in equation (1.4), which is the same whether one uses GLS or OLS, is not the same in more complex multilevel models and so the issue of efficiency arises. Finally, we show how nested projections give rise to intermediate-level omitted variable tests that are useful in model specification. Section 4 outlines extensions and alternatives of this basic approach. We then investigate statistical properties of the estimators and tests in finite samples through a simulation study in Section 5. Section 6 provides an illustration on how to implement the omitted variable tests. Section 7 summarizes the methodology, discusses future directions, and considers topics related to handling omitted variables, and more generally, dealing with unobserved heterogeneity in a population. Appendix A shows how the omitted variable strategy can be employed for models with an arbitrary numbers of levels. Although the notation is complex, the methods can be programmed for routine analysis to provide more flexibility in multilevel modeling. Kim and Frees (2005) provides detailed proofs of properties in the more general linear mixed-effects model case.

2. Three-Level Model Specification and Robust Estimators

In this section, we introduce the three-level model and describe explicitly the type of omitted variables that we consider in this paper. We also provide regression coefficient estimators that are robust to omitted variables. For notation, bold-face symbols represent vectors and matrices whereas scalar quantities are in italics.

2.1 Model Specification

Consider three levels of nesting, where the subscript $s$ identifies a school, the subscript $c$ identifies a child within school $s$, and the subscript $t$ denotes time.

The level-1 model is then written as:

$$ y_{s,c,t} = Z_{s,c,t}^{(1)} \beta_{s,c}^{(1)} + X_{s,c,t}^{(1)} \beta_1 + \epsilon_{s,c,t}^{(1)}, $$

where the variable $y_{s,c,t}$ denotes the response variable (achievement test score); the explanatory variables are $Z_{s,c,t}^{(1)}$ and $X_{s,c,t}^{(1)}$, and may be related to time (grade level, year, etc.), the child (gender, family income, etc.), or school (organization, structure, location, etc.). The parameters that relate to either school $s$ or child $c$ appear as part of the $\beta_{s,c}^{(1)}$ vector, whereas parameters that are constant appear in the $\beta_1$ vector. The disturbance term $\epsilon_{s,c,t}^{(1)}$ has mean zero and variance that is constant over all children and schools. As
presented, this model is assumed not corrupted by omitted variables, except possibly through higher-level terms $\beta^{(1)}_{s,c}$.

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

$$\beta^{(1)}_{s,c} = Z^{(2)}_{s,c} \beta^{(2)} + X^{(2)}_{s,c} \beta^{(2)}_2 + u_{s,c}^{(2)} + \epsilon^{(2)}_{s,c}.$$  

Analogous to the level-1 model, the explanatory variables $Z^{(2)}_{s,c}$ and $X^{(2)}_{s,c}$ relate to the child or school and do not vary over time. The parameters associated with $Z^{(2)}_{s,c}$, $\beta^{(2)}$, may depend on the school $s$ whereas the parameters associated with $X^{(2)}_{s,c}$, $\beta^{(2)}_2$, are constant. The disturbance term $\epsilon^{(2)}_{s,c}$ has mean zero and variance that is constant over children and schools.

Now, let $u^{(2)}_{s,c}$ represent omitted characteristics of the child. Unlike the disturbances $\epsilon^{(1)}_{s,c,f}$ and $\epsilon^{(2)}_{s,c}$, the concern is that the omitted variables $u^{(2)}_{s,c}$ may represent fixed effects or random effects that are correlated with the explanatory variables. If these types of effects are present, then the model does not satisfy the usual assumptions needed for unbiased and consistent regression coefficient estimators.

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

$$\beta^{(2)}_s = X^{(3)}_s \beta_3 + u^{(3)}_s + \epsilon^{(3)}_s.$$  

The variables $X^{(3)}_s$ may depend on the school. The disturbance term $\epsilon^{(3)}_s$ has mean zero and variance that is constant over schools. As with the level-2 model, we let $u^{(3)}_s$ represent omitted characteristics of the school.

### 2.2. Robust Estimation in the Presence of Omitted Variables

To remove the omitted effects $u^{(2)}_{s,c}$ and $u^{(3)}_s$ at both levels 2 and 3, we transform the model. The transformations we propose are not one-to-one and there is a loss of information when applying them. However, they do allow us to develop estimators that are robust to the presence of the omitted effects.

To begin, we combine the three levels into a single equation, linear mixed model as follows:

$$y^{(1)}_{s,c,f} = X^{(1)}_{s,c,f} \beta_1 + \epsilon^{(1)}_{s,c,f} + Z^{(1)}_{s,c,f} \beta^{(1)}_{s,c}$$

$$= X^{(1)}_{s,c,f} \beta_1 + \epsilon^{(1)}_{s,c,f} + Z^{(1)}_{s,c,f} \left( Z^{(2)}_{s,c} \left( X^{(3)}_s \beta_3 + u^{(3)}_s + \epsilon^{(3)}_s \right) + X^{(2)}_{s,c} \beta^{(2)}_2 + u^{(2)}_{s,c} + \epsilon^{(2)}_{s,c} \right).$$

Now, define $Z^{(2)}_{s,c,f} = Z^{(1)}_{s,c,f}$ and $Z^{(3)}_{s,c,f} = Z^{(1)}_{s,c,f} Z^{(2)}_{s,c}$. With this notation, we have

$$y^{(1)}_{s,c,f} = X^{(1)}_{s,c,f} \beta_1 + \epsilon^{(1)}_{s,c,f} + Z^{(3)}_{s,c,f} \left( X^{(3)}_s \beta_3 + u^{(3)}_s + \epsilon^{(3)}_s \right) + Z^{(2)}_{s,c,f} \left( X^{(2)}_{s,c} \beta^{(2)}_2 + u^{(2)}_{s,c} + \epsilon^{(2)}_{s,c} \right)$$

and its stacked version

$$y_{s,c} = X^{(1)}_{s,c} \beta_1 + \epsilon^{(1)}_{s,c} + Z^{(3)}_{s,c} \left( X^{(3)}_s \beta_3 + u^{(3)}_s + \epsilon^{(3)}_s \right) + Z^{(2)}_{s,c} \left( X^{(2)}_{s,c} \beta^{(2)}_2 + u^{(2)}_{s,c} + \epsilon^{(2)}_{s,c} \right).$$
Now, consider a transformation matrix $Q^{(1)}_{s,c}$ that is orthogonal to $Z_{2,s,c}$ so that $Q^{(1)}_{s,c}Z_{2,s,c} = 0$.

Because $Z_{3,s,c} = Z_{2,s,c}Z_{2,s,c}^{(2)}$, we also have that $Q^{(1)}_{s,c}Z_{3,s,c} = 0$. Thus, applying $Q^{(1)}_{s,c}$ to equation (2.2) yields

$$Q^{(1)}_{s,c}y_{s,c} = Q^{(1)}_{s,c}x_{s,c}^{(1)\theta} + Q^{(1)}_{s,c}\varepsilon^{(1)}_{s,c}.$$  

With this transformation we have removed the unobserved, omitted effects.

Using weighted least squares, the transformed model yields the robust estimator

$$b_{1,FE} = \left( X'Q^{(1)}_{s,c}W^{(1)}Q^{(1)}_{s,c}X_{1}\right)^{-1}X'Q^{(1)}_{s,c}W^{(1)}Q^{(1)}_{s,c}y,$$  

where $Q^{(1)} = blkdiag(Q^{(1)}_{s,c})$, $X_{1}$ is the stacked version of $X^{(1)}_{s,c}$ and $W^{(1)}$ is the corresponding matrix of weights. To illustrate, for the OLS version of $b_{1,FE}$, we choose the identity matrix for $W^{(1)}$ and

$$Q^{(1)}_{s,c} = I_{s,c} - Z_{2,s,c}\left(Z_{2,s,c}'Z_{2,s,c}\right)^{-1}Z_{2,s,c}'Z_{2,s,c}.$$  

Alternatively, for the GLS version of $b_{1,FE}$, we choose $W^{(1)} = blkdiag(W^{(1)}_{s,c})$, $W^{(1)}_{s,c} = R^{-1}_{s,c} = (\text{Var}\varepsilon^{(1)}_{s,c})^{-1}$ and

$$Q^{(1)}_{s,c} = I_{s,c} - Z_{2,s,c}\left(Z_{2,s,c}'R_{s,c}^{-1}Z_{2,s,c}\right)^{-1}Z_{2,s,c}'R_{s,c}^{-1}.$$  

As shown in Kim and Frees (2005), $b_{1,FE}$ is an unbiased and asymptotically normal estimator of $\beta_{1}$.

Although complex in appearance, it is easy to compute with standard statistical software. For example, in SAS PROC MIXED, one simply specifies the variables $Z_{2,s,c}$ as categorical factors that will be swept out in the estimation procedure. That is, the estimators can be computed by specifying variables as categorical factors as if one were to assume that $\beta^{(1)}_{s,c}$ were fixed, unknown parameters to be estimated. For this reason, it is customary to refer to the robust estimators as “fixed effects” estimators – we also adopt this terminology. To emphasize this and to distinguish between the different choices of weights, in Sections 5 and 6 we refer to the OLS version of $b_{1,FE}$ as $b_{FC\varepsilon}$ for “fixed child effect estimator.”

For the GLS version, the standard practice is to use a “feasible” GLS estimator, computed in two stages. In the first stage, one uses OLS to compute regression coefficient estimators and residuals. The residuals are then used to estimate variance components that in turn are used to produce feasible GLS estimators in the second stage. We note that it is not always possible to compute a feasible GLS estimator because all of the variance components in $R_{s,c}$ may not be identifiable for a given transformed equation. Kiefer (1980) pointed this out by observing that identification of variance components for an unstructured form of $R$ is impossible. However, the identifiability assumption is viable for standard parametric forms of $R$, such as an $AR(1)$ representation, under mild additional conditions concerning the number of temporal observations.
The estimator $\mathbf{b}_{1,FE}$ is robust to omitted child and school effects. Alternatively, there may be situations in which an analyst is willing to assume that no major child effects are omitted, so that $\mathbf{u}_{s,c} = \mathbf{0}$, but is still concerned with potential omitted school effects. To this end, consider a transformation matrix $Q_s^{(2)}$ so that $Q_s^{(2)} \mathbf{Z}_{3,s} = \mathbf{0}$. Applying this matrix to a stacked version of equation (2.2) yields

$$Q_s^{(2)} \mathbf{y}_s = Q_s^{(2)} \left( \mathbf{X}_s^{(1)} \mathbf{b}_1 + \mathbf{Z}_{s,c} \mathbf{X}_s^{(2)} \mathbf{b}_2 \right) + \delta_s^{(2)},$$

(2.5)

where we use $\delta_s^{(2)} = Q_s^{(2)} \mathbf{e}_s^{(1)} + Q_s^{(2)} \mathbf{Z}_{s,c,} \mathbf{e}_s^{(2)}$ for the disturbance components. Thus, again the transform serves to sweep out the unobserved, omitted effects $\mathbf{u}$. Thus, we use the notation $\mathbf{b}_{2,FE}$ for the corresponding robust estimator of $\mathbf{b}_2$, computed using weighted least squares.

As before, the two most important choices of weights are the identity matrix, corresponding to OLS, and the inverse of the variance of disturbances, corresponding to GLS. In Sections 5 and 6, we refer to the OLS version of $\mathbf{b}_{2,FE}$ as $\mathbf{b}_{FsEE}$ for “fixed school effect estimator” and the GLS version as $\mathbf{b}_{FsRcE}$ for “fixed school, random child effect;” this notation emphasizes how easy it is to compute the robust estimator using a categorical factor for schools. We conjecture that using GLS instead of OLS may be even more important here, because there are now two sources of uncertainty, $\mathbf{e}_1$ and $\mathbf{e}_2$, to account for.

2.3. Special Case: Three-Level Random Intercept Model

As an illustration, we consider the special case of a random intercept model. In this model, there are no slopes associated with the random coefficients so that $\mathbf{Z}_s^{(1)} = \mathbf{1}$ and $\mathbf{Z}_s^{(2)} = \mathbf{1}$. Thus, equation (2.1) yields

$$\mathbf{y}_{s,c} = X_s^{(3)} \mathbf{b}_3 + X_s^{(2)} \mathbf{b}_2 + X_{s,c} \mathbf{b}_1 + \mathbf{e}_s^{(3)} + \mathbf{e}_s^{(2)} + \mathbf{e}_{s,c} + \mathbf{u}_{s,c} + \mathbf{u}_s^{(3)},$$

(2.6)

or, as in equation (2.2),

$$\mathbf{y}_{s,c} = X_s^{(3)} \mathbf{b}_3 + X_s^{(2)} \mathbf{b}_2 + X_{s,c} \mathbf{b}_1 + \mathbf{e}_s^{(3)} + \mathbf{e}_s^{(2)} + \mathbf{e}_{s,c} + \mathbf{u}_{s,c} + \mathbf{u}_s^{(3)}.$$  

For the OLS version, the transformation matrix that removes both child and school omitted effects is

$$Q_{s,c}^{(1)} = \mathbf{I}_{s,c} - \frac{1}{n_{s,c}} \mathbf{1}_{s,c} \mathbf{1}_{s,c}',$$

where $\mathbf{1}_{s,c}$ is a vector of ones having dimension $n_{s,c} \times 1$. (Note that $\mathbf{1}_{s,c} \mathbf{1}_{s,c} = n_{s,c} \mathbf{1}_{s,c}$.)

To interpret the transformed equation (2.3), note that the left hand side is

$$Q_{s,c}^{(1)} \mathbf{y}_{s,c} = \mathbf{y}_{s,c} - \frac{1}{n_{s,c}} \mathbf{1}_{s,c} \mathbf{1}_{s,c}' \mathbf{y}_{s,c} = \mathbf{y}_{s,c} - \mathbf{1}_{s,c} \bar{\mathbf{y}}_{s,c} = \begin{pmatrix} y_{s,c,1} - \bar{y}_{s,c} \\ y_{s,c,2} - \bar{y}_{s,c} \\ \vdots \\ y_{s,c,n_{s,c}} - \bar{y}_{s,c} \end{pmatrix}.$$
because $\frac{1}{n_{x,c}} \mathbf{1}'_{x,c} \mathbf{y}_{x,c} = \frac{1}{n_{x,c}} \sum_{t=1}^{n_{x,c}} \mathbf{y}_{x,c,t} = \overline{\mathbf{y}}_{x,c}$. Thus, the transformed vector of responses is simply the vector of deviations from the time-series averages. To form this directly, take averages over time in equation (2.6) to get

$$\overline{\mathbf{y}}_{x,c} = \mathbf{X}^{(3)} \mathbf{b}_3 + \mathbf{X}^{(2)}_{x,c} \mathbf{b}_2 + \overline{\mathbf{X}}^{(1)}_{x,c} \mathbf{b}_1 + \mathbf{e}^{(3)} + \mathbf{e}^{(2)}_{x,c} + \mathbf{e}^{(1)}_{x,c} + u_{x,c}^{(2)} + u_{x,c}^{(3)}.$$

Subtracting this from equation (2.6) yields

$$\mathbf{y}_{x,c,t} - \overline{\mathbf{y}}_{x,c} = \left( \mathbf{X}^{(1)}_{x,c,t} - \overline{\mathbf{X}}^{(1)}_{x,c} \right) \mathbf{b}_1 + \mathbf{e}^{(1)}_{x,c,t} - \mathbf{e}^{(1)}_{x,c}.$$

This expression for the transformed equation (2.3) shows how the unobserved, omitted effects have dropped out of the equation. Based on these deviations, we have removed the effects of the unobserved variables. The estimator of $\mathbf{b}_1$,

$$\mathbf{b}_{1,FE} = \left( \sum_{x,c,t} \left( \mathbf{X}^{(1)}_{x,c,t} - \overline{\mathbf{X}}^{(1)}_{x,c} \right) \right)^{-1} \left( \sum_{x,c,t} \left( \mathbf{X}^{(1)}_{x,c,t} - \overline{\mathbf{X}}^{(1)}_{x,c} \right) \right)^{\prime} \left( \mathbf{y}_{x,c,t} - \overline{\mathbf{y}}_{x,c} \right),$$

is not corrupted by $\mathbf{u}^{(2)}_{x,c}$ and $\mathbf{u}^{(3)}_{x,c}$. Note that we assume that none of the columns in $\mathbf{X}^{(1)}_{x,c,t}$ is constant over time; otherwise, the deviation from the time-series mean is a constant zero. The more general case that includes time-constant variables is taken up in Section 5.

Similarly, taking averages over time and children in equation (2.6), we have

$$\overline{\mathbf{y}}_{x} = \mathbf{X}^{(3)} \mathbf{b}_3 + \overline{\mathbf{X}}^{(2)} \mathbf{b}_2 + \overline{\mathbf{X}}^{(1)} \mathbf{b}_1 + \mathbf{e}^{(3)} + \mathbf{e}^{(2)} + \mathbf{e}^{(1)} + u^{(2)} + u^{(3)}.$$

Subtracting this from equation (2.6) yields

$$\mathbf{y}_{x,c,t} - \overline{\mathbf{y}}_{x} = \left( \mathbf{X}^{(2)}_{x,c,t} - \overline{\mathbf{X}}^{(2)}_{x,c} \right) \mathbf{b}_2 + \left( \mathbf{X}^{(1)}_{x,c,t} - \overline{\mathbf{X}}^{(1)}_{x,c} \right) \mathbf{b}_1 + \mathbf{e}^{(2)} + \mathbf{e}^{(1)} + u^{(2)} + \left( \mathbf{e}^{(2)} + \mathbf{e}^{(1)} + \mathbf{u}^{(2)} \right).$$

From this equation, we will be able to use fixed effects estimates to get unbiased and consistent estimates of $\mathbf{b}_2$, but only if there are no child-level omitted variables $\mathbf{u}^{(2)}_{x,c}$. In this way we can protect ourselves from school-level omitted variables $\mathbf{u}^{(3)}_{x,c}$. If the analyst makes the assumption that $\mathbf{u}^{(2)}_{x,c} = 0$, then there are two advantages. The first is that additional parameters ($\mathbf{b}_2$) are estimable; the second is that the parameter estimators are more efficient. Of course, making this assumption is an advantage only if it is correct.

### 3. Omitted Variable Tests in Three-Level Models

This section describes two types of omitted variable tests. The first, in Section 3.1, compares a robust estimator to an estimator that is efficient assuming no omitted variables. The second, in Section 3.2, compares robust estimators at different levels.

#### 3.1. Single- and Multiple-Level Omitted Variable Tests

As demonstrated in Section 2.2, it is straightforward to write the three-level model as a single equation, linear mixed effects model. Assuming that there are no omitted variables, the GLS estimators are
easy to compute. The omitted variable test compares a robust estimator to the corresponding GLS estimator. The GLS estimator is efficient, assuming no omitted variables, whereas the robust estimator has desirable properties (such as unbiasedness and asymptotic normality) even in the presence of omitted variables. Because of the efficiency, an analyst would choose to use the GLS estimator if confident that no important variables were omitted. In contrast, the robust estimator would be preferable if there was a concern that important variables were likely to be omitted. The omitted variable test provides valuable guidance regarding selection of the appropriate estimator based on the data.

To see how the test works in a specific situation, we consider estimating the regression coefficients $\beta_1$. If we adopt a hypothesis of no omitted effects, i.e.,
$$H_0: \mathbf{u}^{(2)}_{s,c} = 0 \text{ and } \mathbf{u}^{(3)}_s = 0, \quad \text{for all } s, c,$$
then the random effects estimator, using GLS, is readily available. We will denote it by $b_{1,\text{REE}}$. Alternatively, we also see that a robust estimator, such as $b_{1,\text{FE}}$, is a desirable estimator. One can either use the OLS or GLS version (of the transformed equation) in computing the robust estimator. The test statistic is
$$\chi^2_{\text{FE}} = (b_{1,\text{FE}} - b_{1,\text{REE}})'(\text{Var}b_{1,\text{FE}} - \text{Var}b_{1,\text{REE}})^{-1}(b_{1,\text{FE}} - b_{1,\text{REE}}),$$
where $b_{1,\text{FE}}$ is defined in equation (2.4). Assume that the null hypothesis of no omitted effects in equation (3.1) holds. Then, we show that the statistic $\chi^2_{\text{FE}}$ has an asymptotic chi-square distribution. The degrees of freedom equals the number of parameters in $\beta_1$. Because this is based on testing for omitted variables in two levels in equation (3.1), we refer to it as a multiple-level test.

The two-level, random intercept version of this test has proven popular in the econometrics literature for two important reasons. First, it is intuitive and easy to explain to analysts; the test summarizes the difference between two estimators, both of which are consistent under the null hypothesis and one of which ($b_{1,\text{REE}}$) is not under the alternative. Thus, large values of the test statistic imply rejection of the null hypothesis. Second, both asymptotic variances are standard output from statistical software packages. Thus, computation of the test statistic requires little programming effort. The classic work of Hausman was developed in a (two-level) random intercept model, using only the OLS of transformed models for robust estimation. Properties of the test statistic $\chi^2_{\text{FE}}$ can be described as a special case of our Appendix A results, and thus we defer further discussion to that section.

Assuming no child effects, $\mathbf{u}^{(2)}_{s,c} = 0$, one can also use an omitted variable test to examine the null hypothesis
$$H_0: \mathbf{u}^{(3)}_s = 0, \quad \text{for all } s,$$
of no omitted school effects. This is done by comparing a robust estimator, $b_{2,FE}$, of the regression coefficients $\beta$ to the corresponding GLS estimator, $b_{2,REE}$. The test statistic is

$$\left( b_{2,FE} - b_{2,REE} \right) \left( \text{Var} \left( b_{2,FE} - b_{2,REE} \right) \right)^{-1} \left( b_{2,FE} - b_{2,REE} \right),$$

where $b_{2,FE}$ is defined immediately below equation (2.5). With equation (3.3), this test statistic has an asymptotic chi-square distribution with the degrees of freedom equal to the number of parameters in $\beta$.

Because this is based on testing for omitted variables in one level in equation (3.3), we refer to it as a single-level test.

**Section 3.2 Intermediate-Level Omitted Variable Tests**

It may also be of interest to test for the presence of omitted child effect variables, in the presence of omitted school effect variables. Symbolically, we wish to consider a hypothesis of the form

$$H_0: u_{s,ce}^{(3)} = 0 \quad \text{for all } s, c$$

but allow for the possibility that $u_{s}^{(3)} \neq 0$. Equation (3.4) expresses an assumption about omitted variables at a single level yet allows for the presence of omitted variables at a higher level (level 3); thus, we refer to this as an intermediate-level hypothesis of no omitted variables.

Because we are concerned with potential school-level omitted variables, it is no longer possible to use the random effects estimator. However, we do have two robust estimators available. These are:

- the WLS estimator of the transformed equation (2.3); it is robust to both child and omitted school effects. We denote this by $b_{1,FE(1)}$.
- the GLS estimator of the transformed equation (2.5). It is robust to school omitted effects. We use only the regression coefficients corresponding to $\beta_1$ and denote the estimates of these coefficients by $b_{1,FE(2)}$.

We now introduce an intermediate-level test statistic, defined as

$$\chi^2_{FE(1)} = \left( b_{1,FE(1)} - b_{1,FE(2)} \right) \left( \text{Var} \left( b_{1,FE(1)} - b_{1,FE(2)} \right) \right)^{-1} \left( b_{1,FE(1)} - b_{1,FE(2)} \right),$$

which, with equation (3.4), has an asymptotic chi-square distribution with the degrees of freedom equal to the number of parameters in $\beta_1$.

**4. Extensions**

Sections 2 and 3 introduced the proposed approach in the context of a three-level model. This section extends this basic approach to handle other situations that may be of interest.

The proposed approach extends readily to four and higher order multilevel models; the details of this extension are in Appendix A. This widens the potential scope of applications. Further, the distinction in Section 3 between single, multiple and intermediate level tests becomes much more important in the context of higher order multilevel models.
Appendix A, with further details in Kim and Frees (2005), provide versions of the test statistics using empirical, or “sandwich,” variance estimators. It is customary in mixed linear models to provide hypothesis tests that are robust to heteroscedasticity and serial correlation misspecifications; see, for example, Diggle, Heagarty, Liang, and Zeger (2002), Laird (2004) or Frees (2004). Raudenbush and Bryk (2002) and Maas and Hox (2004) compared asymptotic standard errors calculated by the maximum likelihood method to the empirical standard errors as a way of appraising the possible effect of model misspecification. Arellano (1993) developed Hausman tests for omitted variables in the two-level random intercept model using empirical variance estimators. This extension is particularly important in omitted variable testing because one is concerned with the potential for correlation between disturbance terms and explanatory variables. Thus, it seems sensible to use test statistics that are robust to potential variance component misspecifications. To illustrate, in equation (3.2) one uses an empirical estimate for the variance term \( \left( \text{Var} \beta_{i,FE} - \text{Var} \beta_{i,REE} \right) \). This estimator is a longitudinal data extension of the well-known heteroscedastic-consistent estimator in the linear regression case (Frees, 2004, p.52).

Our asymptotic theory also establishes that the difference in regression coefficient estimators is asymptotically normal. To illustrate, under the null hypothesis in equation (3.1), we have that \( \left( \text{Var} \beta_{i,FE} - \text{Var} \beta_{i,REE} \right)^{-1/2} \left( \beta_{i,FE} - \beta_{i,REE} \right) \) is asymptotically (multivariate) standard normal. This result allows one to examine the impact of each coefficient on the (joint) null hypothesis. As we will see in the Section 6 data example, there are times when such an examination of individual coefficients provides an analyst with insight as to which of the model variables are most susceptible to omitted effects.

We also note that the “sweeping-out” approach for developing robust estimators and test statistics can be modified to handle alternative (more specific) types of omitted variables. To illustrate, one possibility proposed by Ahn, Lee and Schmidt (2001) is to allow omitted effects to vary over time but in a manner that is common to all subjects. Specifically, they investigated the variation of equation (1.2)\[ y_{it} = \alpha_i + \beta_{i} x_{it} + \theta_t u_{it} + \epsilon_{it}, \quad (1.2)^* \]
where \( \theta_t \) is a parameter common to each subject \( i \) yet varies over \( t \). An interpretation offered by Ahn et al. was that one could view \( \theta_t u_{it} \) as subject \( i \)'s reaction to a common macroeconomic shock at time \( t \). As pointed out by Ahn et al., an important advantage of this approach is that time-constant variables are not “swept-out” in the estimation process.

Another possibility, proposed by Verbeke, Spiessens and Lesaffre (2001), is to assume that omitted effects are part of the subject-specific error term but not part of subject-specific slopes. For example, consider the special case where \( Z_{s,c}^{(2)} \) is an identity matrix so that equation (2.1) can be written as
\[ y^{(1)}_{s,c,d} = X^{(1)}_{s,c,d} \beta_1 + \epsilon^{(1)}_{s,c,d} + Z^{(1)}_{s,c,d} \left( X^{(2)}_{s,d} \beta_2 + X^{(3)}_{s} \beta_3 + u^{(2)}_{s,c} + \epsilon^{(2)}_{s,c} + \epsilon^{(3)}_{s} \right). \quad (2.1)^* \]
Now, assume that the unobserved omitted term is of the form \( \mathbf{u}_{s,c}^{(2)} = \left( u_{s,c}^{(2)}, 0, \ldots, 0 \right)^T \) and that the first column of \( \mathbf{Z}_{s,c,t}^{(1)} \) is a constant one, so that \( \mathbf{Z}_{s,c,t}^{(1)} \mathbf{u}_{s,c}^{(2)} = \mathbf{u}_{s,c}^{(2)} \). Similarly, assume that \( \mathbf{u}_{s,c}^{(3)} = \left( u_{s,c}^{(3)}, 0, \ldots, 0 \right)^T \) so that we can write

\[
y_{s,c,t} = \mathbf{X}_{s,c,t}^{(1)} \mathbf{b}_1 + e_{s,c,t}^{(1)} + \mathbf{Z}_{s,c,t}^{(1)} \left( \mathbf{X}_{s,c,t}^{(2)} \mathbf{b}_2 + \mathbf{X}_{s,c,t}^{(3)} \mathbf{b}_3 + e_{s,c,t}^{(2)} + e_{s,c,t}^{(3)} \right) + u_{s,c}^{(2)} + u_{s,c}^{(3)}.
\]

Verbeke et al. (2001) described observational clinical studies where \( t = 1 \) corresponds to baseline measurements and the explanatory variables are interpreted in terms of changes from the baseline. Because of the nature of the sampling scheme, it is intuitively plausible to think of unobserved characteristics affecting baseline measurements. Further, they argued that certain subject-specific changes, as due to exogenous characteristics such as age or time, could be safely be assumed to be unaffected by these omitted variables. For this situation, they established that a mean difference transform (such as in equation 1.3) that removes omitted effects while retaining additional random effects was a more efficient procedure than sweeping out all random effects. In our three-level context, a mean difference transform results in the expression

\[
y_{s,c,t} - \bar{y}_{s,c} = \left( \mathbf{X}_{s,c,t}^{(1)} - \overline{\mathbf{X}}_{s,c}^{(1)} \right) \mathbf{b}_1 + e_{s,c,t}^{(1)} + \left( \mathbf{Z}_{s,c,t}^{(1)} - \overline{\mathbf{Z}}_{s,c}^{(1)} \right) \left( \mathbf{X}_{s,c,t}^{(2)} \mathbf{b}_2 + \mathbf{X}_{s,c,t}^{(3)} \mathbf{b}_3 + e_{s,c,t}^{(2)} + e_{s,c,t}^{(3)} \right).
\]

This transform removes the effects of omitted variables at the child and school level yet allows one to estimate at least some child and school effects.

For both alternatives, knowledge of the way in which omitted effects enters the modeling system should result in better estimators (better as measured by bias and efficiency). In particular, compared to the Verbeke et al. procedure, the section 2 and 3 procedures of sweeping out all random effects are less efficient. The advantages of these procedures are they allow investigators to investigate nested effects at different levels of the hierarchy and do not require special assumptions about the way in which the omitted effects enter a level.

5. Simulation Study

This section investigates the finite sample performance of the robust estimators and omitted variable tests presented in this paper. Data were generated and research design features were manipulated so as to enable greater insight into the types of omitted effects and the corresponding omitted variable tests.

5.1. Data Generation

The foundation for this simulation study is the three-level school-child-time model considered in Sections 2 and 3. To simulate omitted variable effects, we generated a school-level omitted variable \( u_{s}^{(3)} \) and a child-level omitted variable \( u_{s,c}^{(2)} \), three predictors \( X_{s}^{(3)} \), \( X_{s,c}^{(2)} \), and \( \tilde{X}_{s,c,t}^{(1)} \), and three random components \( e_{s}^{(3)} \), \( e_{s,c}^{(2)} \), and \( e_{s,c,t}^{(1)} \) from the standard normal distribution. To allow for correlations between the omitted variables and predictors, we set \( \tilde{X}_{s,c,t}^{(1)} = \tilde{X}_{s,c,t}^{(1)} + \eta_{s} u_{s}^{(3)} + \eta_{s,c} u_{s,c}^{(2)} \). For each potential omitted
variable, the corresponding coefficient $\eta$ was set to two. The outcome variable was generated as a function of predictors, omitted variables, and random components such that

$$y_{s,c,t} = X_{s,c,t}^{(1)} \beta_1 + X_{s,c,t}^{(2)} \beta_2 + X_{s,c,t}^{(3)} \beta_3 + \gamma_{s,c} u_{s,c}^{(3)} + \gamma_{s} u_{s}^{(2)} + \epsilon_{s,c}^{(3)} + \epsilon_{s,c}^{(2)} + \epsilon_{s,c}^{(1)}.$$  

The size of omitted effect was manipulated through the parameter $\gamma$, which was set at levels of 0, 0.5, 1, and 2. The $\beta$ coefficients were fixed at two. Three different combinations of sample sizes were considered for the number of schools ($n_S$), the number of children per school ($n_C$), and the number of time points per child ($n_T$); the levels of ($n_S$, $n_C$, $n_T$) were (10, 10, 5), (20, 10, 5), and (25, 15, 5). Three variance components $\text{Var} \epsilon_{s,c}^{(1)}$, $\text{Var} \epsilon_{s,c}^{(2)}$, and $\text{Var} \epsilon_{s,c}^{(3)}$ were set to one.

Four designs were considered to represent different types of omitted effects: (1) no omitted effects, (2) school effect only ($u_{s,c}^{(3)}$), (3) child effect only ($u_{s,c}^{(2)}$), and (4) both school and child effects ($u_{s,c}^{(3)}$ and $u_{s,c}^{(2)}$), as shown in Table 1. Depending on $\gamma$, we can examine the type I error and power performances of the three hypothesis tests. Recall that the test of omitted school effects requires no omitted child effect, but not vice versa. For each set of simulation conditions, 1,000 datasets were generated.

[Table 1 about here]

### 5.2 Estimators and Tests

We consider four estimators of $\beta_1$ that are compared using three omitted variable tests. The three hypotheses are: $H_0$: $u_{s,c}^{(2)} = u_{s}^{(3)} = 0$ (no omitted effects), $H_0$: $u_{s,c}^{(2)} = 0$ (no omitted child effect), and $H_0$: $u_{s,c}^{(3)} = 0$ (no omitted school effect). In terms of the estimators, the fixed child effects estimator is obtained by the transformation $Q_{s,c}^{(1)}$ in equation (2.3) and is robust to the presence of both $u_{s,c}^{(2)}$ and $u_{s}^{(3)}$. We denote this estimator as $b_{FcEE}$. Using the transformation $Q_{s}^{(2)}$ in equation (2.5), we obtain a fixed school effects estimator, denoted as $b_{FsEE}$, that is robust to $u_{s}^{(3)}$ but not to $u_{s,c}^{(2)}$. The random effects estimator, $b_{REE}$, is most efficient when $u_{s,c}^{(2)} = u_{s}^{(3)} = 0$, but is not robust to either $u_{s}^{(3)}$ or $u_{s,c}^{(2)}$. In addition to the two fixed effects estimators and one random effects estimator, we also consider the fixed school effects and random child effects estimator, denoted as $b_{FsrE}$. Three different tests can be conducted through comparing various pairs of $b_{FcEE}$, $b_{FsrE}$, $b_{FsrE}$, and $b_{REE}$. Specifically, to test $H_0$: $u_{s,c}^{(2)} = u_{s}^{(3)} = 0$, we compare the most robust estimator $b_{FcEE}$ with the most efficient estimator $b_{REE}$. To test $H_0$: $u_{s,c}^{(2)} = 0$, we compare $b_{FcEE}$ with $b_{FsrE}$. To test $H_0$: $u_{s}^{(3)} = 0$, either $b_{FsrE}$ or $b_{FsrE}$ can be used as the robust counterpart of $b_{REE}$. No prior theory supports a preference for one over the other in finite samples.
5.3 Simulation Results

For the performance of the estimators, we investigate the bias and root mean square error (RMSE) of the estimators. These are defined to be:

\[
\text{bias}(b) = \frac{1}{R} \sum_{r=1}^{R} (b_r - \beta) \quad \text{and} \quad \text{RMSE}(b) = \sqrt{\frac{1}{R} \sum_{r=1}^{R} (b_r - \beta)^2},
\]

where \( R = 1,000 \) is the number of replications. For consistency, we report only the summary statistics corresponding to \( \beta_1 \), the level-1 regression coefficient.

To evaluate the performance of the tests, we report the proportion of correct and incorrect rejections. The proportion of rejections is determined as:

\[
\frac{1}{R} \sum_{r=1}^{R} I(\text{test statistic at } r^{th} \text{ simulation exceeds critical value}),
\]

where \( I(.) \) is the indicator function.

Table 2 summarizes the bias and RMSE for the four estimators as a function of \( \gamma \), but averaged across the different sample size conditions, as it was found that the omitted effect size was a more dominant factor than sample size in general. When there is no omitted effect, all estimators are unbiased and their RMSEs are similar except for \( b_{FeEE} \), which has on average 28% larger RMSE than the others. The fixed child effects estimator, \( b_{FcEE} \), is unbiased across all conditions and should be used when the effects of \( u_s^{(3)} \) and \( u_{s,c}^{(2)} \) are both present. Although both \( b_{FeEE} \) and \( b_{FsRE} \) are biased in the presence of \( u_{s,c}^{(2)} \), the amount of bias and magnitude of RMSE in \( b_{FeEE} \) was substantially and consistently larger than for \( b_{FsRE} \). It is conjectured that the poor performance of \( b_{FeEE} \) is due in large part to the incorrect specification of the child level variance component. As \( b_{FsRE} \) is superior to \( b_{FeEE} \) across all conditions, we used \( b_{FsRE} \) instead of \( b_{FeEE} \) for power evaluation in this section and empirical data analysis in the next section. The random effects estimator \( b_{REE} \) is unbiased only when no effect is omitted as the theory suggests. However, \( b_{REE} \) is surprisingly robust to \( u_s^{(3)} \) (the bias was 0.016 when \( \gamma = 2 \)), while being more sensitive to \( u_{s,c}^{(2)} \) (the bias was 0.203 when \( \gamma = 2 \)). When both \( u_s^{(3)} \) and \( u_{s,c}^{(2)} \) are present, the average bias in \( b_{REE} \) was 0.218 when \( \gamma = 2 \), close to the sum of bias associated with each component. Across the conditions, the pattern of bias and RMSE in \( b_{FsRE} \) and \( b_{REE} \) were very similar, although the latter has slightly larger bias.

Although not reported in Table 2, a brief investigation of the RMSE with respect to the three variance components \( \text{Var} \varepsilon_s^{(1)} \), \( \text{Var} \varepsilon_s^{(2)} \), and \( \text{Var} \varepsilon_s^{(3)} \), was also conducted. Here, variance components were estimated via maximum likelihood, under the usual model of no omitted effects. When there is no omitted effect, the average RMSE for the time-, child- and school-level variance components were 0.052, 0.133, and
0.406, respectively, indicating that variability is larger for the higher levels where the number of observations are smaller. There appear to be three factors that contribute to the RMSE of the variance components. The most important factor was the location of omitted effects. When the omission occurs at the school level, the RMSE of school variance increased from 0.406 to 2.017 on average. When the omission occurs at the child level, the RMSE of child variance increased from 0.133 to 1.082 on average. Second, analogous to the estimator, bias and RMSE in variance components increased as a function of omitted effect size. The third factor seems to be sample size. The RMSE was found to be inversely proportional to the number of observations at that level for the sample sizes considered in the simulation. For $\gamma = 0.5$, the RMSE of school variance was 0.725 with $nS = 10$ and 0.503 with $nS = 20$, holding the other conditions constant.

Table 3 summarizes the average rejection rates for the three omitted variable tests. As in Table 2, Table 3 represents the mean of results for different $\gamma$, averaged across sample size. The first and second rows in each condition show results based on model-based variance estimation and empirical variance estimation, respectively. When there is no omitted effect, all are either close to (Tests 1 and 2) or smaller than (Test 3) the normative alpha level of 0.05. The empirical variance estimation statistics tend to be rather conservative compared to their model-based counterparts. When the omitted child effect is present, the test of $u^{(2)}_{s,c} = 0$ has very high power even with $\gamma = 0.5$. The joint occurrence of $u^{(2)}_{s,c}$ and $u^{(3)}_s$ increases the power but only minimally. As noted earlier, Test 3 is not valid when $u^{(2)}_{s,c}$ exists. On the other hand, Test 2 is not affected by $u^{(3)}_s$. Therefore, when only an omitted school effect exists, the rejection rate for Test 2 corresponds to the type I error rate, which was always found to be no larger than 0.05. The chance of detecting the omitted school effect was substantially lower than detecting the omitted child effect, but this does not necessarily imply a lower power for Test 3 but rather rightly reflects the findings in Table 2 in that $u^{(2)}_{s,c}$ yields much larger bias in the estimators than $u^{(3)}_s$ does. Comparing the multiple-level test (Test 1) and the single-level test (Test 3), Test 3 showed higher power than Test 1 in most conditions. The difference between model-based variance estimation and empirical variance estimation were more noticeable in Test 3 than the other tests – for example, 0.910 versus 0.626 when $\gamma = 2$. However, as shown in Figure 1, the sizable difference was found only when the sample size was small ($nS = 10$) and the empirical variance estimation provided comparable power otherwise.

Figure 1 exhibits more detailed results of empirical variance estimation for a subset of Tables 2 and 3. The left upper panel shows the bias in $b_{FBRUE}$ as a function of the omitted child effect and the right upper panel shows the bias in $b_{REE}$ as a function of the omitted school effect. The left lower panel and the right
lower panel depict the rejection rates for Test 2 \((b_{F;EE} \text{ versus } b_{F;RCE})\) and Test 3 \((b_{F;EE} \text{ versus } b_{REE})\), respectively. Both bias and power were increased as the effect size increased, although the likelihood of rejection changed more rapidly for the omitted child effect. Except for one occasion where the effect was small and the sample size was small, the power of Test 2 was close to one. In contrast, the bias in \(b_{REE}\) was moderate even with the large effect, and in this case sample size was more important for improving power. The power also increased as a function of the effect size. In Table 3, the power based on empirical variance estimation was found to be substantially lower than for the model-based variance estimation, but the lower right panel shows that poor performance occurs when the sample size is small (0.11 when \(\gamma =2, N=500\)). The empirical variance estimation provided good power when the sample size was medium or large (0.84 when \(N=1,000\) and 0.95 when \(N=1,875\)).

\[\text{Table 3 and Figure 1 about here}\]

To summarize, the simulation study confirms that the fixed effects estimators are robust to omitted effects; \(b_{F;EE}\) is robust to both \(u^{(3)}_s\) and \(u^{(2)}_{s,c}\), and \(b_{F;EE}\) and \(b_{F;RCE}\) are robust to \(u^{(3)}_s\). All four estimators are unbiased when there is no omitted effect. We found that \(b_{F;RCE}\) is consistently superior to \(b_{F;EE}\), and should be preferred for testing \(u^{(3)}_s = 0\), suggesting that the additional random child intercept in \(b_{F;RCE}\) effectively dilutes bias due to the \(u^{(2)}_{s,c}\) misspecification. The simulation results suggest that an omitted child effect yields more serious problems than an omitted school effect, as \(b_{REE}\) is more sensitive to \(u^{(2)}_{s,c}\) than \(u^{(3)}_s\). The type I error rates are satisfactory for the three tests across all conditions. The omitted effect size was a more important factor in increasing power, but the number of observations also mattered, especially when the bias was not large. The single-level omitted school effect test provides substantially higher power than the multiple-level test in most conditions. The omitted child effect can be tested independent of the omitted school effect, and the intermediate-level test showed low type I error rates and high power with or without \(u^{(3)}_s\). We considered both model-based variance estimation and empirical variance estimation in the simulation study, although the datasets were generated without violation of the model assumptions. It is therefore not surprising that model-based variance estimation showed higher power. Rather, it is encouraging to find that empirical variance estimation provide comparable power except for conditions with a small number of schools \((nS=10)\).

6. Student Achievement Example

Unlike the simulation study, with real data it is impossible to completely quantify the impact of omitted variables in a model of observable data. However, if theory or prior literature leads us to believe that important information is not available for analysis, we can use the estimators and tests presented in this
paper to evaluate the severity of bias due to the omitted effects. This section demonstrates how these statistical methods can be used and interpreted in empirical studies.

The field of school effectiveness has thrived during the past two decades. Among others, Aitkin and Longford (1986), Goldstein and colleagues (Goldstein et al., 1993; Goldstein, 1997; Goldstein & Woodhouse, 2000) and Raudenbush and Willms (1995), and Snijders and Bosker (1999) have considered the assessment of school effectiveness using multilevel models. In particular, the three-level “school-classroom-pupil” and “school-child-time” models are standard configurations in educational research.

6.1. Grade Retention and Switching Schools

This example examines the impact of variables in influencing achievement with special attention given to the potential problem of omitted variables. Among other predictor effects, we wanted to estimate the effects of grade retention and switching schools, two issues of steadfast interest among educational researchers and policy makers (National Association of School Psychologists (NASP), 2003; Dunn, Kadane, & Garrow, 2003). A large number of prior studies exist on the topics of retention and mobility, but empirical research on these issues often yields inconclusive results, partly due to small sample sizes and the analytical difficulties of separating other confounding factors. Previous research appears to suggest that more boys are retained than girls and more minority students are retained than Caucasian students. Large family size, low parental education, and low family involvement are also found to be related to retention (Anderson, Whipple, & Jimerson, 2002). With respect to switching schools, which are closely associated with family events (e.g., parent changing jobs, divorce), high mobility rates have been found particularly harmful for students from low income families in inner city schools (Hanushek, Kain, & Rivkin, 2004).

Different aspects of retention and mobility can be examined and, with longitudinal data, our analysis focuses on intraindividual change related to the two variables, rather than interindividual differences. For example, while most retention research compares outcomes for students who were retained and matched comparisons students who were promoted (Anderson et al., 2002), we examined the effects of retention within a person over time. Also, instead of comparing school performance between high and low turnover rates, we examined changes in test scores for each student after switching schools.

These studies suggest many variables that may be confounded with the effects of retention and switching schools. We considered several such, including student gender, ethnicity, and school information such as the percentage of students who participated in the federal free or reduced-price lunch program, percentage of minority students, class size, and the average years of experience for teachers. Despite our effort to control for confounding effects, however, not all variables that are believed to be important were available in our data. In particular, no variable was available regarding school financial resources (e.g., average teacher salary, per-pupil spending), family income, parent education, or parent occupation. Therefore, our question was whether these unobservable variables would hamper our estimation of the
effects of retention and switching schools, even after we control for gender, ethnicity, and aspects of school quality.

6.2 Data and a Standard Three-Level Model

Webb et al. (2002) examined student achievement on a statewide mathematics test between 1994 and 2000 in Texas, through a study supported by the National Science Foundation Systemic Initiatives Program. As the dataset is extremely large, only a portion of the data was used for the current analysis. We randomly sampled 60 elementary schools in the Dallas public school district, and then sampled 20 students from each school. Test scores were collected annually across four years for each student, producing a total of 2,485 test scores. In the Dallas district, grades three through six all correspond to elementary school. The sampled dataset thus consists of three levels—repeated measures nested within children and children nested within schools. Variables at the three levels are summarized in Table 4. The numbers of test scores per child in the dataset are shown in Table 5.

The level-1 time-varying variables consisted of the dependent variable of mathematics test score, grade, whether a student was retained in the same grade, whether a student switched schools during the past year, and the proportion of students at each school who were eligible for the government free or reduced-price lunch program. Note that this last variable changes over time and so was entered as a level-1 variable. The level-2 child-specific variables were gender, ethnicity, and cohort. The dataset includes 10 cohorts from the elementary class that graduated in 1994 to the class of 2004. Female and Caucasian were used as the reference categories for the corresponding categorical variables. The level-3 school variables were the percentage of minority students, average years of experience of teachers, and average class size.

As a benchmark, we fit a standard three-level random slope and intercept model in which the effect of grade over time was allowed to vary across children. We used GLS, the standard estimation method for multilevel models. The estimates for fixed and random effect parameters and model fit indexes are summarized in Table 6. At level 1, GRADE and RETAINED showed 3.37 and 9.20 point increase on average, respectively. Note that although the mean test score was substantially lower for retained students (59.71 versus 72.43 in Table 4), the coefficient for RETAINED reflects the change within person over time. When students take the test a second time at the same grade level, it might be expected that the estimated effect would be positive. The effects of SWITCHING SCHOOLS and SCHOOL FREE LUNCH RATE were not significant. At level 2, there was no gender difference, but COHORT showed a positive trend over time. African American and Hispanic students on average showed lower scores than Caucasian students even after accounting for the percentage of minority in schools. At level 3, student test scores were negatively correlated with percentage of minority students and positively
correlated with teachers’ experience, but neither was statistically significant. The variance estimates indicated that the unexplained variability was largest at the child level.

[Table 6 about here]

6.3 Omitted Variable Tests

Are the estimated effects of retention and switching schools in Table 6 reliable? Or, are they biased due to the omission of important variables from the analysis? We conducted a test for omitted child and school effects by comparing \( \mathbf{b}_{Fc:EE} \) and \( \mathbf{b}_{REE} \). Recall that only level-1 variables can be estimated in \( \mathbf{b}_{Fc:EE} \), and thus only these were used for the test. Also, the random slope variable GRADE was excluded from the test, resulting in the use of RETAINED, SWITCH SCHOOLS, and SCHOOL FREE LUNCH RATE for the test. The parameter estimates and test results are summarized in the upper part of Table 7. The robust estimators indicate that the RETAINED effect was underestimated and the SWITCHING SCHOOLS effect was overestimated in \( \mathbf{b}_{REE} \) due to omitted effects. Table 7 includes the differences between \( \mathbf{b}_{Fc:EE} \) and \( \mathbf{b}_{REE} \), standard errors for the differences using both model-based variance estimation and empirical variance estimation, the individual coefficient tests, and the “omnibus” test with three degrees of freedom. The model based standard errors are often smaller than the empirical standard errors, but tend to be more sensitive to model misspecification. For robustness, we used empirical standard errors for the individual coefficient tests as well as for the omnibus test. The omnibus test for the null hypothesis \( u_{s,c}^{(2)} = u_{s}^{(3)} = 0 \) was rejected with \( p=0.037 \), and individual coefficient tests indicated that the discrepancy was most severe for RETAINED. Thus, the hypothesis of no omitted effects was rejected, implying that \( \mathbf{b}_{REE} \) is biased. As suspected, the dataset does not include enough information to fit an unbiased random effects model.

[Table 7 about here]

As the multiple level test does not identify the level at which omitted effects occur, we tested \( H_0 : u_{s,c}^{(2)} = 0 \) separately. This intermediate-level test uses the same estimates from \( \mathbf{b}_{Fc:EE} \) as in the multiple-level test, but \( \mathbf{b}_{Fs:RE} \) is the efficient counterpart instead of \( \mathbf{b}_{REE} \). The results are shown at the bottom of Table 7, which shows that the estimates for RETAINED and SWITCH SCHOOLS in \( \mathbf{b}_{Fs:RE} \) were not close to those in \( \mathbf{b}_{Fc:EE} \), and the intermediate-level test was also rejected. This result suggests that there exist omitted child effects and \( \mathbf{b}_{Fs:RE} \) (robust only to the omitted school effects) does not provide unbiased estimates. The rejection of \( u_{s,c}^{(2)} = 0 \) prohibits further testing of \( u_{s}^{(3)} = 0 \), as the validity of the omitted school effect test depends on the assumption of no omitted child effect.

6.4 Summary of Data Analysis

This example examined the effects of grade retention and switching school on student achievement using a longitudinal dataset. Although we controlled for a number of important variables, there was no information collected regarding family income or parent education/occupation, which are theorized to be correlated with
grade retention and mobility. Consequently, the multiple-level and intermediate-level tests showed that the current dataset lacks necessary information to obtain unbiased estimates using efficient estimators. Of course, this result is disappointing in that investigators would like to report results such as in Table 6, that is, estimators for coefficients at all three levels. However, fixed child effects estimators do provide unbiased estimates for retention and mobility, but not for all variables in the model. In addition to the omnibus tests for comparing sets of estimators, we also showed that one can compare a robust estimator and an efficient estimator for each variable. This one degree of freedom test is useful because it helps us to find where the problem is located and distinguishes estimates that are reliable even when the omnibus hypothesis is rejected.

This example demonstrates important implications of omitted variables in school effectiveness research; that is, fixed school effects estimators are not robust to child-level omitted effects and omitted variables at one level may yield biased estimators at any level of the model. Therefore, even though child level variables (e.g., family income, parent education and occupation) may not be of main interest in the study, family characteristics should be collected for the consistent estimation of the effects of other variables that are correlated with child level variables.

As discussed in Section 1.2, when the number of lower level observations (per upper level unit) is large, one can expect fixed and random effects estimators to be similar. Because we have on average $2,485/60 \approx 41$ observations per school, it is not surprising that Table 7 shows that $b_{FsRE}$ estimates are similar to the corresponding $b_{REE}$ estimates. (We thank an anonymous referee for pointing this out to us.) Similarly, with an average of little more than two observations per child, our sampling design provides opportunities to detect important differences in child-level fixed effects and random effects estimators. An alternative design such as the “school-classroom-pupil” design with relatively few observations per school would be more useful for researchers interested in assessing school level omitted effects.

Finally, it should be noted that to employ procedures that provide some protection against omitted variable bias and related issues such as endogeneity, measurement error, and correlated effects involving unobserved variables, analysts must have some knowledge or theory about the source of the omitted variables. ‘Ad-hoc’ searching for additional variables or a ‘data mining’ approach to reduce the bias can make matters worse by exacerbating other biases already present in the data and should be discouraged (Griliches, 1977; Wooldridge 2002). See also our discussion in Section 1.1 on this matter.

7. Summary and Discussion

The omission of important variables is a critical problem in multilevel analysis. As in many observational studies the analyst does not have the ability to collect all the “right” variables; it is of great interest to develop statistical techniques to measure and test the impact of omitted variables and moreover to
provide estimators robust to the presence of omitted variables. This paper presents statistical tools for handling omitted variables in multilevel models.

Whereas most studies involving omitted variables have been limited to two-level panel data models (for exceptions, see Ebbes et al., 2004; McCaffrey, Koretz, Louis, and Hamilton, 2004), this paper explicitly deals with multiple omitted variables at different levels. The simulation study suggests that lower-level omitted variables yield more serious bias in regression coefficients than omitted variables in higher levels with the same degree of correlation with other predictors and the dependent variable. In addition to the conventional fixed-effects estimators and random-effects estimators, this paper dealt with a hybrid of the two approaches and showed its advantageous properties. We also distinguished procedures to test omitted variable effects at multiple, single, and intermediate levels, and have shown how these tests can be used for testing various forms of model misspecification in multilevel models.

In sum, using a general multilevel modeling framework, this study integrates existing approaches into a unified methodology and provides new options for handling omitted variables in multilevel and/or longitudinal data that are ubiquitous in the social and behavioral sciences. It is well known that multilevel models can be written as linear mixed-effects models. However, this paper demonstrates that it can be also helpful to retain the multiple-level representation when inspecting omitted variables at different levels. There remain many issues to be resolved with respect to omitted variables in multilevel models. As for future research, the authors plan to investigate the performance of estimators that are robust to the presence of omitted variables that also retain the ability to account for higher order model relationships by extending Hausman and Taylor’s (1981) instrumental variable estimator to a multilevel framework.

It should be noted that omitted variables can be interpreted as 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 variables, 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., independent error terms, time constant or time varying variables, parametric or nonparametric mixing distributions, etc.) and 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 types, we refer to Heckman and Singer (1982), Chamberlain (1985), Yamaguchi (1986), Palta and Yao (1991), Vermunt (1997), Frank (2000), Halaby (2004), and Frees (2004) for further readings.
References


Rice, N., Jones, A., & Goldstein, H. (1998). Multilevel models where the random effects are correlated with the fixed predictors: a conditioned iterative generalised least squares estimator (CIGLS). York, University of York, Centre for Health Economics.


Appendix A. Higher-Order Multilevel Models

This appendix provides a more precise formulation of the ideas introduced in Sections 2 and 3 in the context of higher-order multilevel models. Appendix A.1 describes higher-order multilevel models, Appendix A.2 provides robust estimators, Appendix A.3 gives single- and multiple-level omitted variable tests and Appendix A.4 describes the intermediate-level omitted variable test.

Appendix A.1 Specification of Higher-order Multilevel Models with Omitted Variables

We adopt the notation system for higher-order multilevel models introduced in Frees and Kim (2006). For a \( k \)-level model, let \( \{i_1, i_2, \ldots, i_k\} = i(k) \) denote the set of \( k \) indices used to identify observations. Here, we might use \( i_k \) as a time index, \( i_{k-1} \) as a child index, \( i_{k-2} \) as a school index, \( i_{k-3} \) as a district index, and so on. We do not wish to restrict consideration to balanced data so we denote the observation set by \( \{i(3): y_{i(3), t} \text{ is observed}\} \). For example, in the context of Section 2.1, the observation set is \( i(3) = \{(s, c, t): y_{s, c, t} \text{ is observed}\} \).

More generally, define \( \{i(k-s): y_{i(k-s), j_k} \text{ is observed for some } j_k\} \) for \( s = 0, 1, \ldots, k-1 \). We will let \( i(k-s) = \{i_1, \ldots, i_{k-s}\} \) be a corresponding element of \( i(k-s) \). For example, \( i(1) \) is the set of all observations that use \( i(1) = i_1 \) for the first index. In the context of Section 2.1, \( i(1) = s \) is the set of all observations that belong to the \( s \)th school. With this additional notation, we are now in a position to provide a recursive specification of higher-order multilevel models.

1. The level-1 model is

\[
y_{i(k)} = \mathbf{Z}_{(i(k))}^{(1)} \beta_{1(i(k))}^{(1)} + \mathbf{X}_{(i(k))}^{(1)} \beta_{1(i(k))}^{(1)} + \epsilon_{(i(k))}^{(1)}, \quad i(k) \in i(k).
\]  

The level-1 fixed parameter vector \( \beta_{1(i(k))} \) has dimension \( K_1 \times 1 \) and the level-1 vector of parameters that may vary over higher levels, \( \beta_{1(i(k))}^{(1)} \), has dimension \( q_1 \times 1 \).

2. For \( g = 2, \ldots, k-1 \), the level-\( g \)-model is

\[
\beta_{g(i(k-1-g))}^{(g-1)} = \mathbf{Z}_{(i(k-1-g))}^{(g)} \beta_{g(i(k-1-g))}^{(g)} + \mathbf{X}_{(i(k-1-g))}^{(g)} \beta_{g(i(k-1-g))}^{(g)} + \mathbf{u}_{(i(k-1-g))}^{(g)} + \epsilon_{(i(k-1-g))}^{(g)}.
\]  

Here, the level-\( g \)-fixed parameter vector \( \beta_{g} \) has dimension \( K_g \times 1 \) and the varying parameter vector \( \beta_{g(i(k-g))}^{(g)} \) has dimension \( q_g \times 1 \). The covariate matrix \( \mathbf{Z}_{(i(k-1-g))}^{(g)} \) has dimension \( q_g \times K_g \) and the predictor matrix \( \mathbf{X}_{(i(k-1-g))}^{(g)} \) has dimension \( q_{g-1} \times K_g \).

3. The level-\( k \)-model is

\[
\beta_{k(i(l))}^{(k-1)} = \mathbf{X}_{(i(l))}^{(k)} \beta_k + \mathbf{u}_{(i(l))}^{(k)} + \epsilon_{(i(l))}^{(k)}.
\]  

(A.3)
Remarks: The quantities \( u_{i(k+1-g)} \) represent unobserved, omitted variables. Note that our procedures do not allow omitted variables in the level-1 model so that we assume \( u_{i(k)} = 0 \). In general, for estimation and testing purposes, we specify an arbitrary level as \( g \) and assume that

\[
\mathbf{u}^{(j)}_{i(k+1-j)} = 0, \quad \text{for } j = 1, \ldots, g. \tag{A.4}
\]

Under this maintained assumption, we will develop estimators that are robust to the presence of omitted variables \( \mathbf{u}^{(j)}_{i(k+1-j)} \), for \( j = g+1, \ldots, k \). Then, we will examine the difference between the robust estimators and the efficient estimators to test for the presence of these omitted variables.

We now write the multiple levels described in equations (A.1)-(A.3) as a single equation, linear mixed effects model suitable for estimation and hypothesis testing. To simplify notation, define the \( 1 \times q_g \) vector

\[
\mathbf{\Gamma} = \mathbf{I} - \mathbf{Z} \mathbf{G} \mathbf{Z}^{-1}, \tag{A.5}
\]

Further, define \( K_g^* = K_1 + \ldots + K_g \), the \( K_g^* \times 1 \) vector \( \mathbf{\beta}_g^* = (\mathbf{\beta}_1', \ldots, \mathbf{\beta}_g')' \) and the \( 1 \times K_g^* \) vector

\[
\mathbf{L} = \mathbf{X} \mathbf{Z} \mathbf{\beta}_g. \tag{A.6}
\]

Now, recursively insert the higher-level models from equations (A.2) and (A.3) into the level-1 equation (A.1). With equation (A.4), we have

\[
\mathbf{y}_{i(k-g)} = \mathbf{\delta}_{i(k-g)} + \mathbf{X}_{g,i(k)} \mathbf{\beta}_g^* + \sum_{j=2}^{g} \mathbf{Z}_{g+1,i(k)} \mathbf{\delta}_{i(k-g)} + \mathbf{Z}_{g+1,i(k)} \mathbf{\beta}_{i(k-g)}^{(j)}, \quad i(k-g) \in i(k-g), \tag{A.6}
\]

where \( \mathbf{\delta}_{i(k-g)} \) is the stacked version of \( \mathbf{\delta}_{i(k)} + \sum_{j=2}^{g} \mathbf{Z}_{j,i(k)} \mathbf{\delta}_{i(k-g)}^{(j)} \). Under the assumption in equation (A.4), we have that \( \mathbf{\delta}_{i(k-g)} \) are i.i.d. over different indices of the form \( \{i(k-g)\} \) and have mean zero. Thus, equation (A.6)
specifies a mixed linear model. The terms $\beta_{i(k-g)}^{(g)}$ include potentially important unobserved, omitted variables of the form $u_{i(k+1-j)}$, for $j = g+1, \ldots, k$. Additional stacking permits us to re-write equation (A.6) as

$$ y = \delta^{(g)} + X_g \beta_g^* + Z_{g+1} \beta^{(g)}. \quad (A.7) $$

Appendix A.2. Robust Estimation in the Presence of Omitted Variables

To remove the omitted effects, we use a transform $Q^{(g)}$ such that $Q^{(g)} Z_{g+1} = 0$. We wish to allow for the possibility that this transform may also zero out certain columns of $X_g$. Thus, we partition $X_g \beta_g^*$ so that

$$ X_g \beta_g^* = \left( X_{g,1} : X_{g,2} \right) \begin{pmatrix} \beta_{g,1}^* \\ \beta_{g,2}^* \end{pmatrix} = X_{g,1} \beta_{g,1}^* + X_{g,2} \beta_{g,2}^* $$

and assume that $X_{g,2}$ is collinear with $Z_{g+1}$ in the sense that $Q^{(g)} X_{g,2} = 0$.

Now, one can apply the transform to equation (A.6) and use weighted least squares to derive estimators of $\beta_{g,1}^*$. Specifically, we define

$$ b_{g,1,FE}^{*,*} = \left( X_{g,1} Q^{(g)} X_{g,1} \right)^{-1} X_{g,1} Q^{(g)} W^{(g)} Q^{(g)} y, \quad (A.8) $$

where $W^{(g)}$ is a matrix of weights. We focus on two important special cases.

For the first case, we use OLS; here, we choose the identity matrix for $W^{(g)}$ and the transform

$$ Q_{i(k-g)}^{(g)} = I_{i(k-g)} - Z_{g+1,i(k-g)} \left( Z_{g+1,i(k-g)}^T Z_{g+1,i(k-g)} \right)^{-1} Z_{g+1,i(k-g)}^T, $$

and define $Q^{(g)} = blkdiag(Q_{i(k-g)}^{(g)})$. For the second case, we use GLS; here, we choose $W^{(g)} = blkdiag(W_{i(k-g)}^{(g)})$, $W_{i(k-g)}^{(g)} = V_{\delta_{i(k-g)}}^{-1}$ and the transform

$$ Q_{i(k-g)}^{(g)} = I_{i(k-g)} - Z_{g+1,i(k-g)} \left( Z_{g+1,i(k-g)} V_{\delta_{i(k-g)}}^{-1} Z_{g+1,i(k-g)} \right)^{-1} Z_{g+1,i(k-g)} V_{\delta_{i(k-g)}}^{-1}. $$

As seen in Kim and Frees (2005), both estimators are unbiased and asymptotically normal estimators of $\beta_{g,1}^*$. If one correctly specifies $V_{\delta_{i(k-g)}}$, then the GLS version is more efficient than the OLS version, even in the presence of omitted variables.

Equation (A.6) is simply a mixed linear model with replicates $i(k-g) \in i(k-g)$, fixed effects $X_{g,i(k-g)} \beta_g^*$, random effects $Z_{g+1,i(k-g)} \beta^{(g)}_{i(k-g)}$, and disturbances $\delta_{i(k-g)}$. Thus, from an extension of Maddala’s (1971) result to mixed linear effects models (see Frees, 2004, Exercise 7.5), it is immediate that we can write
\[ b_{g,1,REE}^* = (I - \Delta^*) b_{g,1,FE}^* + \Delta^* b_{g,1,B}^* , \]

where \( b_{g,1,B}^* \) is a “between-groups” estimator of \( \beta_g^* \), and \( \Delta^* \) is a measure of the relative precision of the estimators. As noted in Section 1.2, we may interpret this result to mean that one can anticipate more prominent differences between \( b_{g,1,REE}^* \) and \( b_{g,1,FE}^* \) when the number of lower level observations per upper level unit is small or the variability at the upper level is small.

**Appendix A.3. Single- and Multiple-Level Omitted Variable Tests**

Now consider the hypothesis of no unobserved, omitted variables; that is,

\[ H_0: u_{(k+1-j)}^{(j)} = 0, \quad \text{for } j = g+1, \ldots, k. \quad (A.9) \]

Under this assumption and \( g=1 \), the model described in Appendix A.1 follows the usual (without omitted effects) multilevel model formulation. Denote \( b_{g,1,REE}^* \) to be the usual (efficient) random effects estimator of \( \beta_g^* \), based on generalized least squares. When \( g+1 = k \), we are testing only the highest level, \( u_{(1)}^{(k)} = 0 \), and we refer to this as a single-level hypothesis. More generally, for \( g+1 < k \), we are testing several levels simultaneously and thus refer to equation (A.9) as a multiple-level hypothesis.

We are now in a position to state the omitted variable test that extends the result of Hausman (1978) to the multilevel framework.

**Proposition A1.** Consider the multilevel model specified in Appendix A.1 and assume that the null hypothesis of no omitted effects in equation (A.9) holds. Then, \( b_{g,1,FE}^* - b_{g,1,REE}^* \) is asymptotically normal with asymptotic variance \( \left( \text{Var} b_{g,1,FE}^* - \text{Var} b_{g,1,REE}^* \right) \) and the statistic

\[ \chi^2_{FE} = \left( b_{g,1,FE}^* - b_{g,1,REE}^* \right)' \left( \text{Var} b_{g,1,FE}^* - \text{Var} b_{g,1,REE}^* \right)^{-1} \left( b_{g,1,FE}^* - b_{g,1,REE}^* \right) \]

has an asymptotic chi-square distribution. The degrees of freedom equals the number of parameters in \( \beta_{g,1}^* \).

**Remarks:** Because the null hypothesis is no omitted effects, when applying the test of omitted variables, if the test statistic \( \chi^2_{FE} \) is large, this provides evidence of important omitted variables. Like the basic Hausman test, the statistic \( \chi^2_{FE} \) is easy to compute because the estimators and the corresponding variance-covariance matrices are output from standard mixed linear model routines and little special programming is required. Unlike the basic Hausman test, we prove the result in Kim and Frees (2005) for robust estimators.
determined by weighted least squares. This allows the analyst to specify alternative versions of robust estimators, such as in the OLS or GLS versions, depending on one’s knowledge of the variance structure.

It is also customary in mixed linear models to provide hypothesis tests that are robust to heteroscedasticity and serial correlation misspecifications. To illustrate, Diggle et al. (2002) and Laird (2004) refer to such tests as using robust, sandwich or empirical variance estimators. In econometrics, this was given by Arellano (1993) for Hausman tests in the two-level random intercept model. Kim and Frees (2005), Property B8, provides the details of the robust version of the test statistic.


An additional hypothesis test of interest is the following. Suppose that the researcher acknowledges omitted variables at levels $g+2, \ldots, k$. For example, in the school-child-time model, the researcher acknowledges that important school variables are omitted (at level $g+2=3$). However, the interest is in testing the hypothesis of no unobserved, omitted variables, at the $(g+1)$st level, that is,

$$H_0: \mathbf{u}_{(k-g)}^{(g+1)} = \mathbf{0}. \quad (A.10)$$

In equation (A.10), one assumes no omitted variables at levels $g$ and below yet allows for omitted variables at levels $g+2$ and above; thus, we refer to this hypothesis as an intermediate-level one. We are able to test this hypothesis using robust estimators based on nested projections, as follows.

To introduce the nested projections, at the $(g+1)$st level we may use a WLS estimator. To illustrate, we use the projection corresponding to OLS,

$$Q_{i(k-g)}^{(g)} = \mathbf{I}_{i(k-g)} - Z_{g+1,i(k-g)} \left( Z_{g+1,i(k-g)} Z_{g+1,i(k-g)} \right)^{-1} Z_{g+1,i(k-g)}.'$$

At the $(g+2)$nd level, the transform for the GLS estimator is

$$Q_{i(k-g-1)}^{(g+1)} = \mathbf{I}_{i(k-g-1)} - Z_{g+2,i(k-g-1)} \left( Z_{g+2,i(k-g-1)} V_{6,i(k-g-1)}^{-1} Z_{g+2,i(k-g-1)} \right)^{-1} Z_{g+2,i(k-g-2)} V_{6,i(k-g-1)}^{-1}.$$

To compare these projections, we define

$$Q_{i(k-g-1)}^{(g)} = \text{blkdiag}(Q_{i(k-g),1}, \ldots, Q_{i(k-g),n(i(k-g))}).$$

With this notation, using equation (A.5), we have

$$Q_{i(k-g-1)}^{(g)} Z_{g+2,i(k-g-1)} = \begin{pmatrix}
Q_{i(k-g),1}^{(g)} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \ddots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & Q_{i(k-g),n(i(k-g))}^{(g)}
\end{pmatrix}
\begin{pmatrix}
Z_{g+1,i(k-g),1}^{(g+1)} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \ddots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & Z_{g+1,i(k-g),n(i(k-g))}^{(g+1)}
\end{pmatrix}
= \begin{pmatrix}
Q_{i(k-g),1}^{(g)} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \ddots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & Q_{i(k-g),n(i(k-g))}^{(g)}
\end{pmatrix} = \mathbf{0}.

Thus,

$$Q_{i(k-g-1)}^{(g)} Q_{i(k-g-1)}^{(g+1)} = Q_{i(k-g-1)}^{(g)}.$$

In this sense, we say that the projections $\left\{Q_{i(k-g)}^{(g)} \right\}_{g=1}^{k-1}$ are nested.
Now, we may apply the \((g+2)nd\) level (GLS) transform to the model in equation (A.7). Using equation (A.2), we have

\[
Q^{(g+1)}y = Q^{(g+1)}\delta^{(g)} + Q^{(g+1)}X_{g}g\beta_{g}^{*} + Q^{(g+1)}Z_{g+1}g\beta_{g}^{(g)}
\]

\[
= Q^{(g+1)}\left(\delta^{(g)} + Z_{g+1}g\epsilon^{(g+1)}\right) + Q^{(g+1)}\left(X_{g}g\beta_{g}^{*} + Z_{g+1}gX^{(g+1)}g\beta_{g+1}^{*}\right) + Q^{(g+1)}Z_{g+1}g\epsilon^{(g+1)}.
\]

By this transformation, we have removed all sources of omitted variables except at the \((g+2)nd\) level. Thus, under the null hypothesis in equation (A.10), we may compute the robust (GLS) estimator. Further, applying the transform at the \((g+1)st\) level, we have

\[
Q^{(g)}y = Q^{(g)}Q^{(g+1)}y = Q^{(g)}\delta^{(g)} + Q^{(g)}X_{g}g\beta_{g}^{*},
\]

because \(Q^{(g)}Q^{(g+1)} = Q^{(g)}\) and \(Q^{(g)}Z_{g+1} = 0\). This transform sweeps out the omitted variables \(u^{(g+1)}\), enabling us to compute an estimator that is robust to their presence.

In summary, the computation of the intermediate-level omitted variable test can be performed as follows:

- Compute the WLS estimator in equation (A.8) at the \((g+1)st\) level. We denote this by \(b_{g,1,FE(g)}^{*}\).
- Compute the GLS estimator in equation (A.8) at the \((g+2)nd\) level. Use only the estimated regression coefficients corresponding to \(\beta_{g}^{*}\); we denote this by \(b_{g,1,FE(g+1)}^{*}\).
- Compute the intermediate-level test statistic

\[
\chi_{FE(g)}^{2} = \left(b_{g,1,FE(g)}^{*} - b_{g,1,FE(g+1)}^{*}\right)^{\top} \left(\text{Var}b_{g,1,FE(g)}^{*} - \text{Var}b_{g,1,FE(g+1)}^{*}\right)^{-1} \left(b_{g,1,FE(g)}^{*} - b_{g,1,FE(g+1)}^{*}\right). \tag{A.11}
\]

Properties of the intermediate-level test statistic are immediate from Proposition A1 and the above discussion. We summarize them in the following:

**Proposition A2.** Consider the multilevel model specified in Appendix A.1 and assume that the null hypothesis of no omitted effects in equation (A.10) holds. Then, \(b_{g,1,FE(g)}^{*} - b_{g,1,FE(g+1)}^{*}\) is asymptotically normal with asymptotic variance \(\left(\text{Var}b_{g,1,FE(g)}^{*} - \text{Var}b_{g,1,FE(g+1)}^{*}\right)\) and the test statistic \(\chi_{FE(g)}^{2}\) has an asymptotic chi-square distribution. The degrees of freedom equals the number of parameters in \(\beta_{g,1}^{*}\).

**Remarks:** Although we have presented the intermediate-level hypothesis based on a single level in equation (A.10), our results also apply directly to multiple levels.
Table 1.
Properties of estimators and tests with four types of omitted effects.

<table>
<thead>
<tr>
<th>Type of Estimator</th>
<th>Property of Estimator</th>
<th>Type of Omitted Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed child effects estimator ($b_{FcEE}$)</td>
<td>unbiased</td>
<td>No Omitted Variable Effect</td>
</tr>
<tr>
<td>Fixed school effects estimator ($b_{FsEE}$)</td>
<td>unbiased</td>
<td>Child Effect ($u_{c}^{(2)}$)</td>
</tr>
<tr>
<td>Fixed school and random child effects estimator ($b_{FsRcE}$)</td>
<td>unbiased</td>
<td>School Effect ($u_{s}^{(3)}$)</td>
</tr>
<tr>
<td>Random effects estimator ($b_{REE}$)</td>
<td>unbiased</td>
<td>Child and School Effects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>($u_{s,c}^{(2)}, u_{s}^{(3)}$)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type of Test</th>
<th>Property of Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Multiple-level test</td>
<td>$H_0: u_{s}^{(3)} = u_{s,c}^{(2)} = 0$</td>
</tr>
<tr>
<td>2. Intermediate-level test</td>
<td>$H_0: u_{s,c}^{(2)} = 0$</td>
</tr>
<tr>
<td>3. Single-level test</td>
<td>$H_0: u_{s}^{(3)} = 0$</td>
</tr>
</tbody>
</table>
Table 2. 
Estimated bias (first row) and RMSE (second row) for $\gamma = 0, 0.5, 1.0, \text{ and } 2.0$, where $\gamma$ is the coefficient of the omitted variable and indicates the strength of the omitted variable effect.

<table>
<thead>
<tr>
<th>Type of Estimator</th>
<th>Type of Omitted Effects</th>
<th>No Omitted Variable Effect</th>
<th>Child Effect $(u_{x,c}^{(2)})$</th>
<th>School Effect $(u_{s}^{(3)})$</th>
<th>Child and School Effects $(u_{x,c}^{(2)}, u_{s}^{(3)})$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma = 0$</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$\gamma = 0.5$</td>
<td>0.38</td>
<td>0.003</td>
<td>0.003</td>
<td>0.037</td>
</tr>
<tr>
<td>Fixed child effects estimator ($b_{FCEE}$)</td>
<td>$\gamma = 1.0$</td>
<td>0.000</td>
<td>1.0</td>
<td>0.001</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>$\gamma = 2.0$</td>
<td>0.000</td>
<td>2.0</td>
<td>0.001</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>$\gamma = 0$</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>Fixed school effects estimator ($b_{FSEE}$)</td>
<td>$\gamma = 0$</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$\gamma = 0.5$</td>
<td>0.196</td>
<td>0.005</td>
<td>0.005</td>
<td>0.197</td>
</tr>
<tr>
<td></td>
<td>$\gamma = 1.0$</td>
<td>0.199</td>
<td>0.002</td>
<td>0.002</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td>$\gamma = 2.0$</td>
<td>0.787</td>
<td>0.001</td>
<td>0.001</td>
<td>0.785</td>
</tr>
<tr>
<td>Fixed school and random child effects estimator ($b_{FsrE}$)</td>
<td>$\gamma = 0$</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$\gamma = 0.5$</td>
<td>0.101</td>
<td>0.011</td>
<td>0.014</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>$\gamma = 1.0$</td>
<td>0.106</td>
<td>0.016</td>
<td>0.014</td>
<td>0.107</td>
</tr>
<tr>
<td></td>
<td>$\gamma = 2.0$</td>
<td>0.183</td>
<td>0.018</td>
<td>0.018</td>
<td>0.183</td>
</tr>
<tr>
<td>Random effects estimator ($b_{REE}$)</td>
<td>$\gamma = 0$</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$\gamma = 0.5$</td>
<td>0.102</td>
<td>0.014</td>
<td>0.014</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>$\gamma = 1.0$</td>
<td>0.107</td>
<td>0.016</td>
<td>0.016</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>$\gamma = 2.0$</td>
<td>0.208</td>
<td>0.018</td>
<td>0.018</td>
<td>0.218</td>
</tr>
</tbody>
</table>
Table 3.
Average rejection rates. The first row is calculated using model-based standard errors and the second row is calculated using empirical standard errors.

<table>
<thead>
<tr>
<th>Type of Estimator</th>
<th>Type of Omitted Effects</th>
<th>No Omitted Variable Effect</th>
<th>Child Effect ((u_{s,c}^{(2)}))</th>
<th>School Effect ((u_{s}^{(3)}))</th>
<th>Child and School Effects ((u_{s,c}^{(2)}, u_{s}^{(3)}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Multiple-level test (H_0: u_{s}^{(3)} = u_{s,c}^{(2)} = 0)</td>
<td>(\gamma = 0)</td>
<td>0.051</td>
<td>0.955</td>
<td>0.230</td>
<td>0.962</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>0.915</td>
<td>0.200</td>
<td>0.926</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.000</td>
<td>1.0</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.000</td>
<td>0.456</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>1.000</td>
<td>0.446</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.000</td>
<td>0.353</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>2. Intermediate-level test (H_0: u_{s,c}^{(2)} = 0)</td>
<td>(\gamma = 0)</td>
<td>0.051</td>
<td>0.954</td>
<td>0.049</td>
<td>0.959</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>0.912</td>
<td>0.040</td>
<td>0.916</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.000</td>
<td>1.0</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.000</td>
<td>0.048</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>1.000</td>
<td>0.050</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.000</td>
<td>0.046</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>3. Single-level test (H_0: u_{s}^{(3)} = 0)</td>
<td>(\gamma = 0)</td>
<td>0.033</td>
<td>0.270</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>0.179</td>
<td>0.179</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>0.663</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.460</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>0.910</td>
<td>0.910</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.626</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.
Variable descriptions with proportions and mean mathematics scores by subgroups.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Percentage</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level-1 variables (replications over time)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRADE</td>
<td>Grade 3 to Grade 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RETAINED</td>
<td>Retained in the same grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>YES</td>
<td>0.56</td>
<td>59.71</td>
</tr>
<tr>
<td></td>
<td>NO</td>
<td>99.44</td>
<td>72.43</td>
</tr>
<tr>
<td>SWITCH SCHOOLS</td>
<td>Switched schools in a particular year</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>YES</td>
<td>6.76</td>
<td>71.76</td>
</tr>
<tr>
<td></td>
<td>NO</td>
<td>93.24</td>
<td>72.41</td>
</tr>
<tr>
<td>SCHOOL FREE LUNCH RATE</td>
<td>The proportion of economically disadvantaged students eligible for the federal free or reduced-price lunch program each year (mean: 0.76).</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Level-2 variables (replications over child)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENDER</td>
<td>MALE</td>
<td>50.92</td>
<td>71.89</td>
</tr>
<tr>
<td></td>
<td>FEMALE</td>
<td>49.08</td>
<td>71.15</td>
</tr>
<tr>
<td>ETHNICITY</td>
<td>AFRICAN AMERICAN</td>
<td>48.00</td>
<td>69.57</td>
</tr>
<tr>
<td></td>
<td>HISPANIC</td>
<td>38.67</td>
<td>72.13</td>
</tr>
<tr>
<td></td>
<td>CAUCASIAN</td>
<td>11.67</td>
<td>82.06</td>
</tr>
<tr>
<td></td>
<td>OTHER (Asian, Native Indian, Mixed, etc.)</td>
<td>1.66</td>
<td>76.08</td>
</tr>
<tr>
<td>COHORT</td>
<td>Elementary school graduating class of 1994 to 2004 (1 to 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Level-3 variables (replications over school)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEACHER EXPERIENCE</td>
<td>Average years of teaching</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(mean: 12.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLASS SIZE</td>
<td>Average class size (mean: 20.73)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PERCENT_MINORITY</td>
<td>Percentage of minority students</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(mean: 94.98)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.
Observed test scores per child.

<table>
<thead>
<tr>
<th>Number of test scores per child</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>41.58</td>
</tr>
<tr>
<td>2</td>
<td>24.67</td>
</tr>
<tr>
<td>3</td>
<td>19.17</td>
</tr>
<tr>
<td>4</td>
<td>14.25</td>
</tr>
<tr>
<td>5</td>
<td>0.33</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100.00</td>
</tr>
</tbody>
</table>
Table 6.
Random effects models with a variable slope for GRADE_3=GRADE-3. *: $p<0.05$, **: $p<0.01$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>$t$-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTERCEPT</td>
<td>69.78</td>
<td>7.82**</td>
</tr>
<tr>
<td>GRADE_3</td>
<td>3.37</td>
<td>17.51**</td>
</tr>
<tr>
<td>RETAINED</td>
<td>9.20</td>
<td>4.20**</td>
</tr>
<tr>
<td>SWITCH SCHOOLS</td>
<td>-0.37</td>
<td>-0.54</td>
</tr>
<tr>
<td>SCHOOL FREE</td>
<td>-0.23</td>
<td>-0.19</td>
</tr>
<tr>
<td>LUNCH RATE</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Level 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENDER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Female</td>
<td>-1.23</td>
<td>-1.52</td>
</tr>
<tr>
<td>ETHNICITY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>-4.74</td>
<td>-2.98**</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-3.61</td>
<td>-2.32*</td>
</tr>
<tr>
<td>Other</td>
<td>6.53</td>
<td>1.93</td>
</tr>
<tr>
<td>Caucasian</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>COHORT</td>
<td>1.50</td>
<td>8.74**</td>
</tr>
<tr>
<td><strong>Level 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEACHER EXPERIENCE</td>
<td>0.16</td>
<td>0.92</td>
</tr>
<tr>
<td>CLASS SIZE</td>
<td>0.07</td>
<td>0.38</td>
</tr>
<tr>
<td>PERCENT_MINORITY</td>
<td>-0.12</td>
<td>-1.57</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance Components</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Var} e_{s,t,t}^{(1)}$</td>
<td>45.33</td>
<td>2.24</td>
</tr>
<tr>
<td>$\text{Var} e_{s,c}^{(2)}$ - Intercept</td>
<td>242.02</td>
<td>15.01</td>
</tr>
<tr>
<td>$\text{Var} e_{s,c}^{(2)}$ - Covariance</td>
<td>24.75</td>
<td>3.83</td>
</tr>
<tr>
<td>$\text{Var} e_{s,c}^{(2)}$ - Slope</td>
<td>4.71</td>
<td>1.34</td>
</tr>
<tr>
<td>$\text{Var} e_{s}^{(3)}$</td>
<td>4.75</td>
<td>2.85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model Fit Indices</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-$2 \log L$</td>
<td>19,143.8</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>19,179.8</td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>19,285.5</td>
<td></td>
</tr>
</tbody>
</table>

- $p<0.05$, **$p<0.01$
Table 7. Tests of omitted child and school effects using empirical standard errors.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$b_{F_{cE}}$</th>
<th>$b_{REE}$</th>
<th>Difference</th>
<th>Model-based standard error</th>
<th>Empirical standard error</th>
<th>Individual coefficient test ($\chi^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RETAINED</td>
<td>14.55</td>
<td>9.20</td>
<td>5.35</td>
<td>1.28</td>
<td>2.48</td>
<td>4.65, $p=0.03$</td>
</tr>
<tr>
<td>SWITCH SCHOOLS</td>
<td>-1.54</td>
<td>-0.37</td>
<td>-1.18</td>
<td>0.62</td>
<td>0.63</td>
<td>3.37, $p=0.06$</td>
</tr>
<tr>
<td>SCHOOL FREE LUNCH RATE</td>
<td>1.99</td>
<td>-0.23</td>
<td>2.22</td>
<td>1.66</td>
<td>1.85</td>
<td>1.44, $p=0.23$</td>
</tr>
</tbody>
</table>

$H_0: u^{(2)}_{s,c} = u^{(3)}_s = 0, \chi^2 = 8.59 (df = 3, p = 0.035)$

Tests of omitted child effects using empirical standard errors.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$b_{F_{cE}}$</th>
<th>$b_{F_{dRE}}$</th>
<th>Difference</th>
<th>Model-based standard error</th>
<th>Empirical standard error</th>
<th>Individual coefficient test ($\chi^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RETAINED</td>
<td>14.55</td>
<td>9.34</td>
<td>5.22</td>
<td>1.28</td>
<td>2.45</td>
<td>4.54, $p=0.03$</td>
</tr>
<tr>
<td>SWITCH SCHOOLS</td>
<td>-1.54</td>
<td>-0.30</td>
<td>-1.24</td>
<td>0.62</td>
<td>0.65</td>
<td>3.70, $p=0.05$</td>
</tr>
<tr>
<td>SCHOOL FREE LUNCH RATE</td>
<td>1.99</td>
<td>0.58</td>
<td>1.41</td>
<td>1.61</td>
<td>1.74</td>
<td>0.66, $p=0.42$</td>
</tr>
</tbody>
</table>

$H_0: u^{(2)}_{s,c} = 0, \chi^2 = 8.47 (df = 3, p = 0.037)$