Journal of Educational and Behavioral Statistics
2014, Vol. 39, No. 6, pp. 524–549
DOI: 10.3102/1076998614559420

Handling Correlations Between Covariates and Random Slopes in Multilevel Models

Michael David Bates
Michigan State University

Katherine E. Castellano
Educational Testing Service

Sophia Rabe-Hesketh
University of California, Berkeley

Anders Skrondal
Norwegian Institute of Public Health

This article discusses estimation of multilevel/hierarchical linear models that include cluster-level random intercepts and random slopes. Viewing the models as structural, the random intercepts and slopes represent the effects of omitted cluster-level covariates that may be correlated with included covariates. The resulting correlations between random effects (intercepts and slopes) and included covariates, which we refer to as ''cluster-level endogeneity,'' lead to bias when using standard random effects (RE) estimators such as (restricted) maximum likelihood. While the problem of correlations between unit-level covariates and random intercepts *is well known and can be handled by fixed-effects (FE) estimators, the problem of correlations between unit-level covariates and random* slopes *is rarely considered. When applied to models with random slopes, the standard FE estimator does not rely on standard cluster-level exogeneity assumptions, but requires an ''uncorrelated variance assumption'' that the variances of unit-level covariates are uncorrelated with their random slopes. We propose a ''per-cluster regression'' (PC) estimator that is straightforward to implement in standard software, and we show analytically that it is unbiased for all regression coefficients under cluster-level endogeneity and violation of the uncorrelated variance assumption. The PC, RE, and an augmented FE estimator are applied to a real data set and evaluated in a simulation study that demonstrates that our PC estimator performs well in practice.*

Keywords: *endogeneity; hierarchical linear model; multilevel model*

1 Introduction

We consider linear regression models for clustered data that include cluster-specific random intercepts and slopes. Such models are called multilevel models, mixed models, random-coefficient models, or hierarchical linear models. If the models are viewed as "structural models," the perspective taken in this article, the regression coefficients represent structural or causal parameters, and the error terms represent the effects of omitted covariates. If there are omitted confounders that are correlated with included covariates, then the error terms are correlated with the included covariates. These correlations lead to omitted variable bias. This article focuses on estimation methods that avoid bias due to omitted cluster-level confounders, also referred to as "cluster-level endogeneity." An alternative view of models, not taken here, is that regression coefficients merely represent associations between included variables, or linear projections in the case of linear models, in which case the error terms are orthogonal to the covariates by construction (see Spanos, 2006, for a discussion of "structural" vs. "statistical" models).

Research on addressing cluster-level endogeneity in multilevel models has traditionally been confined to correlations between unit-level covariates (i.e., covariates that vary over units) and random *intercepts* that vary over clusters in which units are nested. This constitutes a type of "cluster-level" endogeneity, as it involves correlation with a cluster-level random error term. For instance, in estimating the effect of Catholic schooling on student achievement controlling for student socioeconomic status (SES), one may worry that school-level omitted variables, such as school resources, may be correlated with SES. Left unaddressed, this endogenity may lead us to misattribute the impacts of these omitted variables to the effect of SES. This bias may in turn spill over to other coefficients. To address this type of endogeneity, Mundlak (1978) shows that consistent estimators of the coefficients of unit-level covariates can be obtained by a fixed-effects (FE) approach. However, with standard FE estimators, coefficients of cluster-level covariates (i.e., covariates that only vary over clusters) cannot be estimated. The Hausman and Taylor (1981) instrumental variable estimator resolves this limitation and is consistent for the coefficients of both unit- and cluster-level covariates under appropriate assumptions (see Castellano, Rabe-Hesketh, & Skrondal, 2014).

Endogeneity in the form of correlations between unit-level covariates and random *slopes* varying over clusters may also arise. Referring back to the Catholic schooling example, when controlling for students' SES, the slope of SES (or SES achievement gradient) may vary between schools, due to interactions between SES and omitted school-level covariates, such as school resources. If the omitted variables are negatively correlated with the SES achievement gradient, then the random slopes will be negatively correlated with SES.

Remarkably, such endogeneity is rarely considered. One exception is Frees (2004) who extends the Mundlak approach to handle random slopes. Another is Wooldridge (2005) who shows under seemingly benign conditions that traditional FE estimation of random-intercept models is robust against correlations between unit-level covariates and random slopes. However, neither of these approaches permits estimation of the coefficients of cluster-level covariates even if the covariates are exogenous (i.e., not endogenous). This limitation is overcome by Kim and Frees (2007) who use generalized method of moments estimation to extend the Hausman–Taylor approach to multilevel models with random slopes. However, their method is difficult to implement, making the FE approach more feasible in practice.

Unfortunately, a key assumption required for the FE approach may be violated in important applications. Specifically, the within-cluster variance of a unit-level covariate must be uncorrelated with the random slope of that covariate, which we refer to as the ''uncorrelated variance assumption'' throughout this article. Turning back to the Catholic schooling example, it is possible that more diverse schools (schools with high variance of SES) may be better equipped to mitigate the effects of SES (lower the SES achievement gradient) than schools that are more homogeneous (schools with low variance of SES). Such a situation would directly violate the uncorrelated variance assumption.

In this article, we investigate estimation of the coefficients of unit-level and cluster-level covariates in multilevel models in the presence of two sources of endogeneity; (nonzero) correlations between unit-level covariates and *both* the random intercept and random slopes. Throughout, we assume that covariates are uncorrelated with the unit-level error term and that cluster-level covariates are uncorrelated with the random intercepts and random slopes. We propose a simple ''per-cluster regression'' (PC) approach that is unbiased and consistent for coefficients of both unit-level and cluster-level covariates under both forms of cluster-level endogeneity and violation of the uncorrelated variance assumption. We contrast its performance to the ''standard'' random-effects (RE) estimator and what we call the ''augmented FE'' (FE+) approach, which extends the FE approach to provide estimates of the coefficients of cluster-level covariates. In Section 2, we first introduce our motivating empirical example and specific model of interest and then present our general model. In Section 3, we discuss the traditional RE and FE estimators and their conditions for unbiasedness. In Section 4, we introduce new estimators, namely, the FE+ estimator and the PC estimator, and show under what assumptions they are unbiased. We provide conditions for consistency for all four estimators in Online Appendix A. All estimators are applied to a data set in Section 5, and Section 6 investigates performance of the estimators using a simulation study.

2 Motivation and Multilevel Model

2.1 Motivating Example and Specific Model

As a motivating example, we consider the effect of private schooling on student achievement. We use the Raudenbush and Bryk (2002) data from the 1982 High School and Beyond (HSB) Survey because it is familiar in education and it is in the public domain, allowing us to provide data and commands for all estimators in Online Appendix B.

This two-level data set provides us with an estimation sample of 7,185 students (units) nested in 160 schools (clusters), 70 of which are Catholic (private) and the remaining of which are public. The number of observations per school ranges from 14 to 67 students (mean = 45, $SD = 12$). We use a mathematics standardized test score for student i in school j as our response variable, y_{ij}, which has a mean of 12.75 and a standard deviation of 6.88. Our primary variables of interest are w_j, a binary indicator for whether school j is Catholic, and x_{ij}, a continuous index of students' SES, composed of parental education, parental occupation, and parental income. This index has a mean of 0 and a standard deviation of 0.78.

We write the model using a two-stage formulation, similar to that used in Raudenbush and Bryk (2002). The first stage is the Level-1 model:

$$y_{ij} = \eta_{0j} + \eta_{1j} x_{ij} + \varepsilon_{ij}. \tag{1}$$

This is a simple regression of the student mathematics test scores y_{ij} on their SES x_{ij}, where the intercept η_{0j} and slope η_{1j} can vary between schools, as indicated by the j subscript. Each student's test score can deviate from the school-specific regression line by a random error term ε_{ij}.

The school-specific intercepts and slopes become (unobservable) outcomes in the Level-2 models:

$$\begin{aligned} \eta_{0j} &= \gamma_0 + \gamma_1 w_j + u_{0j} \\ \eta_{1j} &= \beta_1 + \beta_2 w_j + u_{1j}. \end{aligned} \tag{2}$$

The mean intercept and slope of SES, for the population of schools, depend on whether the schools are Catholic or public (w_j). The intercepts γ_0 and β_1 in these models therefore represent the population means of the intercepts and slopes of SES for public schools, whereas the slopes γ_1 and β_2 represent the differences in population means of the intercepts and slopes between Catholic and private schools, respectively. The Level-2 models have errors u_{0j} and u_{1j} to allow the schools' intercepts and slopes to vary within the subpopulations of Catholic and private schools. Assumptions regarding the error terms are discussed in Sections 3 and 4.

Substituting the Level-2 models into the Level-1 model, we obtain the reduced form of the model:

$$y_{ij} = \gamma_0 + \gamma_1 w_j + \beta_1 x_{ij} + \beta_2 w_j x_{ij} + u_{0j} + u_{1j} x_{ij} + \varepsilon_{ij}. \tag{3}$$

We see that β_2 is the coefficient of a cross-level interaction between the student-level covariate x_{ij} and the school-level covariate w_j.

In this setting, it is likely that there are omitted school-level variables that affect student achievement, and hence enter the random intercept u_{0j}, and are correlated with student SES. If these omitted school-level variables also interact with student SES, then they enter the random slope u_{1j}, and the slope is then correlated with SES. Ignoring such endogeneity may lead to bias when estimating the coefficients of this model.

2.2 General Multilevel Model

The general model we consider in this article is for two-level data, such as the HSB data described previously. In the cross-sectional case, units (Level 1) are typically individuals nested within clusters (Level 2), such as schools, hospitals, or neighborhoods. In the longitudinal case, units refer to measurement occasions nested within individuals who constitute the "clusters." Clusters are indexed j, with $j = 1, \ldots, J$, and units are indexed ij, with $i = 1, \ldots, n_j$. The general model includes unit-level covariates that vary between units within clusters (and between clusters) as well as cluster-level covariates that vary between clusters but are constant across units within the same cluster. Same-level and cross-level interactions may also be included, where cross-level interactions are unit-level covariates. Some unit-level covariates may have random slopes.

The general model can be written as:

$$\mathbf{y}_i = \mathbf{W}_j\boldsymbol{\gamma} + \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\varepsilon}_j. \tag{4}$$

Here $\mathbf{y}_j = (y_{1j}, \ldots, y_{n_jj})'$ is the vector of responses for cluster j, the $n_j \times (P+1)$ matrix $\mathbf{W}_j$ includes all P cluster-level covariates and its first column is a vector of ones, $\mathbf{1}_{n_j}$, for the intercept; the $n_j \times R$ matrix $\mathbf{X}_j$ includes all R unit-level covariates; the $n_j \times (R_1 + 1)$ matrix $\mathbf{Z}_j$ includes all $R_1 \leq R$ unit-level covariates in $\mathbf{X}_j$ that have random slopes and its first column is $\mathbf{1}_{n_j}$ for the random intercept. Finally, the $\mathbf{u}_j$ are random effects or cluster-level error terms (one random intercept and R_1 random slopes), and the $\boldsymbol{\varepsilon}_j$ are unit-level error terms. The $\mathbf{u}_j$ are assumed to be independent of the $\boldsymbol{\varepsilon}_j$, and the clusters are independent in the sense that error terms as well as covariates are independent across clusters. Other required assumptions regarding $\mathbf{u}_j$ and $\boldsymbol{\varepsilon}_j$ depend on what estimators are used and are discussed in Sections 3 and 4. Sometimes all unit-level covariates have random slopes, so that $R_1 = R$, but typically $R_1 < R$, so that the slopes of some unit-level covariates do not vary between clusters, giving rise to the term "mixed-effects" (mixed random and "fixed" effects) model.

For the specific model of the HSB data in Equation 3, $P = 1$, so that $\mathbf{W}_j$ has n_j rows and $P + 1 = 2$ columns, with each row equal to $(1, w_j)$. The corresponding coefficients are $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)'$. The matrix of $R = 2$ unit-level covariates is

$\mathbf{X}_j = (\mathbf{x}_j\ w_j\mathbf{x}_j)$, where $\boldsymbol{x}_j = \left(x_{1j}, \ldots, x_{n_j j}\right)'$, and $\boldsymbol{\beta} = (\beta_1, \beta_2)'$. The first unit-level covariate has a random slope, so $R_1 = 1$, $\mathbf{Z}_j = \left(\mathbf{1}_{n_j}\ \boldsymbol{x}_j\right)$, and $\mathbf{u}_j = \left(u_{0j}, u_{1j}\right)'$. Finally, the unit-level errors are $\boldsymbol{\varepsilon}_j = \left(\varepsilon_{1j}, \ldots, \varepsilon_{n_j j}\right)'$.

The model in Equation 4 can be expressed more compactly by combining all covariates (i.e., $\mathbf{X}_j$ and $\mathbf{W}_j$) into a single matrix $\mathbf{V}_j$ and likewise their corresponding coefficients ($\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$) into a single vector $\boldsymbol{\delta}$:

$$\mathbf{y}_j = \mathbf{V}_j\boldsymbol{\delta} + \mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\varepsilon}_j. \tag{5}$$

It will often be convenient to stack the covariates for all J clusters in the matrix $\mathbf{V}$.

Writing this general model using a two-stage formulation, analogous to Equations 1 and 2, requires additional notation, so we defer this until Section 4.2 where this formulation is useful for explaining the PC approach.

3 Standard Estimators and Conditions for Unbiasedness

3.1 Exogeneity and Endogeneity

Throughout this article, we assume unit-level exogeneity or strict exogeneity, given the random effects,

$$\mathrm{E}\left(\boldsymbol{\varepsilon}_j|\mathbf{V}_j, \mathbf{u}_j\right) = \mathbf{0}. \tag{6}$$

It follows that $\mathrm{E}\left(\boldsymbol{\varepsilon}_j|\mathbf{V}_j\right) = \mathbf{0}$ and $\mathrm{E}\left(\boldsymbol{\varepsilon}_j'\mathbf{V}_j\right) = \mathbf{0}$ and that each element of $\boldsymbol{\varepsilon}_j$ is uncorrelated with each element of the covariate vectors.

The assumption of cluster-level exogeneity can be expressed as:

$$\mathrm{E}\left(\mathbf{u}_j|\mathbf{V}_j\right) = \mathbf{0}. \tag{7}$$

When this assumption is violated, there is cluster-level endogeneity.

We assume that cluster-level exogeneity holds for the cluster-level covariates

$$\mathrm{E}\left(\mathbf{u}_j|\mathbf{W}_j\right) = \mathbf{0} \tag{8}$$

and discuss the problem that unit-level covariates may be cluster-level endogenous,

$$\mathrm{E}\left(\mathbf{u}_j|\mathbf{X}_j\right) \neq \mathbf{0}.$$

Cluster-level endogeneity occurs if, for example, $\mathrm{E}\left(\mathbf{X}_j'\mathbf{1}_{n_j}u_{0j}\right) \neq \mathbf{0}$ or in other words, if the cluster sums or means of the unit-level covariates are correlated with the random intercepts.

In this article, we consider RE, FE+, and (our proposed) PC approaches for estimating the model in Equation 3 when $\mathbf{X}_j$ is correlated with both the random intercept u_{0j} and the random slopes u_{rj}, $r = 1, \ldots, R_1$. In the following

subsections, we describe each of these estimators and under which conditions they are unbiased for the regression coefficients $\boldsymbol{\delta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$.

3.2 RE estimators

For RE estimators, the random effects, $\mathbf{u}_j$, are assumed to have zero means and covariance matrix $\boldsymbol{\Psi}$, given the covariates. They are uncorrelated across clusters, and they are also uncorrelated with the unit-level error term $\boldsymbol{\varepsilon}_j$. The elements ε_{ij} of $\boldsymbol{\varepsilon}_j$ have zero means and variance θ, given the covariates, and are mutually uncorrelated.

It follows from these assumptions that the mean and covariance structure of $\mathbf{y}_j$, given $\mathbf{V}_j$ becomes

$$\mathrm{E}\left(\mathbf{y}_j|\mathbf{V}_j\right) = \mathbf{V}_j\boldsymbol{\delta},$$

and

$$\boldsymbol{\Sigma}_j \equiv \mathrm{Var}\left(\mathbf{y}_j|\mathbf{V}_j\right) = \mathbf{Z}_j\boldsymbol{\Psi}\mathbf{Z}_j' + \theta\mathbf{I}_{\mathrm{n_j}}. \tag{9}$$

Under the exogeneity assumptions in Equations 6 and 7, consistent estimators for the parameters of the model in Equation 5 can be obtained using maximum likelihood (ML), restricted ML (REML), or feasible generalized least squares (FGLS). The FGLS estimator can be expressed in closed form as

$$\widehat{\boldsymbol{\delta}}_{\mathrm{RE}} = \boldsymbol{\delta} + \left(J^{-1}\sum_{j=1}^{J}\mathbf{V}_j'\widehat{\boldsymbol{\Sigma}}_j^{-1}\mathbf{V}_j\right)^{-1}\left(J^{-1}\sum_{j=1}^{J}\mathbf{V}_j'\widehat{\boldsymbol{\Sigma}}_j^{-1}\left(\mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\varepsilon}_j\right)\right), \tag{10}$$

where $\widehat{\boldsymbol{\Sigma}}_j$ is an estimator of $\boldsymbol{\Sigma}_j$, obtained by substituting estimators of $\boldsymbol{\Psi}$ and θ into Equation 9, and $\sum_j \mathbf{V}_j'\widehat{\boldsymbol{\Sigma}}_j^{-1}\mathbf{V}_j$ is assumed to be nonsingular with probability 1.

The conditional expectation of the GLS estimator $\widehat{\boldsymbol{\delta}}_{\mathrm{GLS}}$ (assuming $\boldsymbol{\Sigma}$ is known), given $\mathbf{V}$, is

$$\mathrm{E}\left(\widehat{\boldsymbol{\delta}}_{\mathrm{GLS}}|\mathbf{V}\right) = \boldsymbol{\delta} + \left(J^{-1}\sum_{j=1}^{J}\mathbf{V}_j'\boldsymbol{\Sigma}_j^{-1}\mathbf{V}_j\right)^{-1}\left(J^{-1}\sum_{j=1}^{J}\mathbf{V}_j'\boldsymbol{\Sigma}_j^{-1}\left[\mathbf{Z}_j\mathrm{E}\left(\mathbf{u_j}|\mathbf{V}\right) + \mathrm{E}\left(\boldsymbol{\varepsilon}_j|\mathbf{V}\right)\right]\right).$$

Note that $\mathrm{E}\left(\boldsymbol{\varepsilon}_j|\mathbf{V}\right) = \mathrm{E}\left(\boldsymbol{\varepsilon}_j|\mathbf{V}_j\right)$ and $\mathrm{E}\left(\mathbf{u}_j|\mathbf{V}\right) = \mathrm{E}\left(\mathbf{u}_j|\mathbf{V}_j\right)$ because clusters are assumed to be independent. Furthermore, unit-level exogeneity implies that $\mathrm{E}\left(\boldsymbol{\varepsilon}_j|\mathbf{V}_j\right) = \mathbf{0}$ and cluster-level exogeneity implies that $\mathrm{E}\left(\mathbf{u}_j|\mathbf{V}_j\right) = \mathbf{0}$. Conditional unbiasedness, $\mathrm{E}\left(\widehat{\boldsymbol{\delta}}_{\mathrm{GLS}}|\mathbf{V}\right) = \boldsymbol{\delta}$, hence follows; and using the law of iterated expectations, $\widehat{\boldsymbol{\delta}}_{\mathrm{GLS}}$ is (unconditionally) unbiased, $\mathrm{E}\left(\widehat{\boldsymbol{\delta}}_{\mathrm{GLS}}\right) = \boldsymbol{\delta}$.

Unfortunately, due to the nonlinear nature of the FGLS estimator, this unbiasedness result does not automatically apply when estimates $\widehat{\boldsymbol{\Sigma}}_j$ are plugged in for

$\boldsymbol{\Sigma}_j$. Under the previous assumptions, a sufficient assumption for unbiasedness, $\mathrm{E}\left(\widehat{\boldsymbol{\delta}}_{\mathrm{FGLS}}\right) = \boldsymbol{\delta}$, is that the error terms have symmetric distributions (Kakwani, 1967). This result also applies to ML and REML, given that these estimators can be expressed as iterative versions of the FGLS estimator (Don & Magnus, 1980). Note that for the empirical example and simulation study, we use REML, following the tradition of its use in the education research literature.

3.3 FE Estimators

In econometrics, the term fixed-effects (FE) estimator refers to an estimator that does not rely on cluster-level exogeneity, and we adopt this terminology here. Some of the estimators can be derived by treating the random effects as fixed and others by eliminating the random effects, but in either case, the effects are typically viewed as random. The traditional FE approaches have been developed for random-intercept models, with $\mathbf{Z}_j = \mathbf{1}_{n_j}$ and $\mathbf{u}_j = u_{0j}$, to handle violation of the exogeneity assumption $\mathrm{E}\left(u_{0j}|\mathbf{V}_j\right) = 0$ (Mundlak, 1978).

We define the FE estimator in terms of the de-meaning (or group-mean centering) transformation $\mathbf{Q}_j \equiv \mathbf{I}_{n_j} - \mathbf{1}_{n_j}\left(\mathbf{1}_{n_j}'\mathbf{1}_{n_j}\right)^{-1}\mathbf{1}_{n_j}'$ where $\mathbf{Q}_j\mathbf{1}_{n_j} = \mathbf{0}$, and define $\ddot{\mathbf{y}}_j \equiv \mathbf{Q}_j\mathbf{y}_j$, $\ddot{\mathbf{X}}_j \equiv \mathbf{Q}_j\mathbf{X}_j$, $\ddot{\mathbf{Z}}_j \equiv \mathbf{Q}_j\mathbf{Z}_j$, and $\ddot{\boldsymbol{\varepsilon}}_j \equiv \mathbf{Q}_j\boldsymbol{\varepsilon}_j$. Premultiplying Equation 4 by $\mathbf{Q}_j$, the de-meaned model becomes

$$\ddot{\mathbf{y}}_j = \ddot{\mathbf{X}}_j\boldsymbol{\beta} + \ddot{\mathbf{Z}}_j\mathbf{u}_j + \ddot{\boldsymbol{\varepsilon}}_j.$$

Note that $\ddot{\mathbf{W}}_j\boldsymbol{\gamma} = \mathbf{0}_{n_j}$ because the columns of $\mathbf{W}_j$ are constant. The first column of $\ddot{\mathbf{Z}}_j$ is $\mathbf{0}_{n_j}$ because the first column of $\mathbf{Z}_j$ is $\mathbf{1}_{n_j}$, so that the random part of the model does not depend on u_{0_j}. The FE estimator can be obtained by applying a pooled ordinary least squares (POLS) estimator (pooling over the Level-2 units) to the de-meaned model, giving

$$\widehat{\boldsymbol{\beta}}_{\mathrm{FE}} = \boldsymbol{\beta} + \left(J^{-1}\sum_{j=1}^{J}\ddot{\mathbf{X}}_j'\ddot{\mathbf{X}}_j\right)^{-1}\left(J^{-1}\sum_{j=1}^{J}\ddot{\mathbf{X}}_j'\left(\ddot{\mathbf{Z}}_j\mathbf{u}_j + \ddot{\boldsymbol{\varepsilon}}_j\right)\right),$$

where $\sum_j \ddot{\mathbf{X}}_j'\ddot{\mathbf{X}}_j$ is assumed to be nonsingular with probability 1.

To derive the conditions for unbiasedness, stack the $\ddot{\mathbf{X}}_j$ for all clusters j in $\ddot{\mathbf{X}}$. Consider the conditional expectation of $\hat{\boldsymbol{\beta}}_{\mathrm{FE}}$, given $\ddot{\mathbf{X}}$,

$$\mathrm{E}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}|\ddot{\mathbf{X}}\right) = \boldsymbol{\beta} + \left(J^{-1}\sum_{j=1}^{J}\ddot{\mathbf{X}}_j'\ddot{\mathbf{X}}_j\right)^{-1}\left(J^{-1}\sum_{j=1}^{J}\ddot{\mathbf{X}}_j'\left[\ddot{\mathbf{Z}}_j\mathrm{E}\left(\mathbf{u}_j|\ddot{\mathbf{X}}\right) + \mathrm{E}\left(\ddot{\boldsymbol{\varepsilon}}_j|\ddot{\mathbf{X}}\right)\right]\right). \quad (11)$$

Keep in mind that $\ddot{\mathbf{Z}}_j$ is a subset of $\ddot{\mathbf{X}}_j$ and note that $\mathrm{E}\left(\mathbf{u}_j|\ddot{\mathbf{X}}\right) = \mathrm{E}\left(\mathbf{u}_j|\ddot{\mathbf{X}}_j\right)$ and $\mathrm{E}\left(\ddot{\boldsymbol{\varepsilon}}_j|\ddot{\mathbf{X}}\right) = \mathrm{E}\left(\ddot{\boldsymbol{\varepsilon}}_j|\ddot{\mathbf{X}}_j\right)$ because clusters are independent. Furthermore, unit-level

exogeneity implies that $\mathrm{E}(\ddot{\boldsymbol{\varepsilon}}|\ddot{\mathbf{X}}_j) = \mathbf{0}$, and cluster-level exogeneity implies that $\mathrm{E}(\mathbf{u}_j|\ddot{\mathbf{X}}_j) = \mathbf{0}$, so $\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}$ is conditionally unbiased, $\mathrm{E}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}|\ddot{\mathbf{X}}\right) = \boldsymbol{\beta}$. Finally, using the law of iterated expectations, it follows that $\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}$ is (unconditionally) unbiased, $\mathrm{E}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}\right) = \boldsymbol{\beta}$.

Wooldridge (2005, 2010, section 11.7.3) considers FE estimation for a special case of Equation 4 without cluster-level covariates $\mathbf{W}_j$. He derives the conditions required for consistency of the traditional FE estimator, and we briefly derive the analogous results for the general model (Equation 4) in Online Appendix A, Section A.2. The condition for consistency $p\lim \widehat{\boldsymbol{\beta}}_{\mathrm{FE}} = \boldsymbol{\beta}$, is

$$\mathrm{E}\left(\ddot{\mathbf{X}}_j'\ddot{\mathbf{Z}}_j\mathbf{u}_j\right) = \mathbf{0}. \tag{12}$$

As pointed out by Wooldridge (2010, p. 382), this assumption allows $\mathbf{u}_j$ to be correlated with the "permanent" components of $\mathbf{X}_j$, but not the "idiosyncratic" components $\ddot{\mathbf{X}}_j$. Condition 12 is implied by the more easily interpretable condition for unbiasedness,

$$\mathrm{E}(\mathbf{u}_j|\ddot{\mathbf{X}}_j) = \mathbf{0},$$

because $\mathrm{E}\left(\ddot{\mathbf{X}}_j'\ddot{\mathbf{Z}}_j\mathbf{u}_j|\ddot{\mathbf{X}}_j\right) = \ddot{\mathbf{X}}_j'\ddot{\mathbf{Z}}_j\mathrm{E}(\mathbf{u}_j|\ddot{\mathbf{X}}_j)$.

We now look at the assumption in Equation 12, which is required for consistency and unbiasedness, in more detail for our specific model for the empirical example in Equation 3 to understand how it can be violated. In that model, $\ddot{\mathbf{X}}_j = \left(\ddot{\mathbf{x}}_j\ w_j\ddot{\mathbf{x}}_j\right)$ and $\ddot{\mathbf{Z}}_j = \left(\mathbf{0}_{n_j}\ \ddot{\mathbf{x}}_j\right)$. The condition can be written as two nontrivial equations,

$$\mathrm{E}\left[\begin{pmatrix}\ddot{\mathbf{x}}_j' \\ w_j\ddot{\mathbf{x}}_j'\end{pmatrix}\ddot{\mathbf{x}}_j u_{1j}\right] = \mathbf{0}.$$

Concentrating on the first equation, we obtain

$$\mathrm{E}\left(\ddot{\mathbf{x}}_j'\ddot{\mathbf{x}}_j u_{1j}\right) = \mathrm{E}\left[\left(\sum_{i=1}^{n_j}\ddot{x}_{ij}^2\right)u_{1j}\right] = (n_j - 1)\mathrm{E}\left(s_j^2 u_{1j}\right) = 0, \tag{13}$$

where s_j^2 is the sample variance of x_{ij} for cluster j. In other words, the condition is violated if the within-cluster variance of x_{ij} is correlated with the random slope u_{1j}. In our empirical example, it seems reasonable that more diverse schools (larger s_j^2) may be better suited to mitigate the effects of SES (smaller u_{1j}) than more homogeneous schools. We will therefore consider the problem of nonzero correlation between the within-cluster variance of x_{ij} and its random slope u_{1j} and will refer to Equation 12 as the "uncorrelated variance assumption." Note, however, that Equation 12 can also be violated in other ways. For instance, in longitudinal data, the covariate value at the initial time point, x_{1j}, may be correlated with u_{1j}.

4 New Estimators and Conditions for Unbiasedness

4.1 FE+ Estimator

The FE approach does not permit estimation of the coefficients, $\boldsymbol{\gamma}$, of any cluster-level covariates, but it can be "augmented" so that it does. The FE+ estimator we use is similar to the estimator proposed by Hausman and Taylor (1981) to estimate effects of cluster-level covariates for two-level models with only random intercepts—and no random slopes. As pointed out by Castellano, Rabe-Hesketh, and Skrondal (2014), the estimator discussed here has been invented and reinvented several times (e.g., Ballou, Sanders, & Wright, 2004; Raudenbush & Willms, 1995; Wiley, 1975). However, the conditions for unbiasedness for random-coefficient models have not previously been considered.

Step 1: Estimation of $\boldsymbol{\beta}$

In the first step, $\boldsymbol{\beta}$ (coefficients for all unit-level covariates) is estimated by FE. The estimator is unbiased under the uncorrelated variance assumption in Equation 12 and the unit-level exogeneity assumption in Equation 6.

Step 2: Estimation of $\boldsymbol{\gamma}$

In the second step, we obtain quasi-residuals r_j,

$$\mathbf{r}_j \equiv \mathbf{y}_j - \mathbf{X}_j\widehat{\boldsymbol{\beta}}_{\mathrm{FE}},$$

and estimate $\boldsymbol{\gamma}$ (coefficients for the cluster-level covariates) in the regression of $\mathbf{r}_j$ on $\mathbf{W}_j$ by POLS. Substituting Equation 4 for $\mathbf{y}_j$, we obtain

$$\mathbf{r}_j = \mathbf{W}_j\boldsymbol{\gamma} + \mathbf{X}_j\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{\mathrm{FE}}\right) + \mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\varepsilon}_j.$$

The POLS estimator of $\boldsymbol{\gamma}$ can be expressed as

$$\widehat{\boldsymbol{\gamma}} = \boldsymbol{\gamma} + \left(J^{-1}\sum_{j=1}^{J}\mathbf{W}_j'\mathbf{W}_j\right)^{-1}\left(J^{-1}\sum_{j=1}^{J}\mathbf{W}_j'\left[\mathbf{X}_j\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{\mathrm{FE}}\right) + \mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\varepsilon}_j\right]\right), \qquad (14)$$

where $\sum_j \mathbf{W}_j'\mathbf{W}_j$ is assumed to be nonsingular with probability 1.

The conditional expectation of the POLS estimator $\widehat{\boldsymbol{\gamma}}$ given $\mathbf{V}$, becomes

$$\mathrm{E}(\widehat{\boldsymbol{\gamma}}|\mathbf{V}) = \boldsymbol{\gamma} + \left(J^{-1}\sum_{j=1}^{J}\mathbf{W}_j'\mathbf{W}_j\right)^{-1}\left(J^{-1}\sum_{j=1}^{J}\mathbf{W}_j'\left\{\mathbf{X}_j\left[\boldsymbol{\beta} - \mathrm{E}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}|\mathbf{V}\right)\right]\right.\right.$$

$$\left.\left. + \mathbf{Z}_j\,\mathrm{E}\left(\mathbf{u}_j|\mathbf{V}\right) + \mathrm{E}\left(\boldsymbol{\varepsilon}_j|\mathbf{V}\right)\right\}\right).$$

Conditional unbiasedness, $\mathrm{E}(\widehat{\boldsymbol{\gamma}}|\mathbf{V}) = \boldsymbol{\gamma}$ follows since (i) $\mathrm{E}(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}|\mathbf{V})] = \boldsymbol{\beta}$ under the uncorrelated variance assumption, (ii) $\mathrm{E}\left(\mathbf{u}_j|\mathbf{V}\right) = \mathrm{E}\left(\mathbf{u}_j|\mathbf{V}_j\right)$ and

$\mathrm{E}(\boldsymbol{\varepsilon}_j|\mathbf{V}) = \mathrm{E}(\boldsymbol{\varepsilon}_j|\mathbf{V}_j)$ due to independence of the clusters, and (iii) $\mathrm{E}(\mathbf{u}_j|\mathbf{V}_j) = \mathbf{0}$ and $\mathrm{E}(\boldsymbol{\varepsilon}_j|\mathbf{V}_j) = \mathbf{0}$ under the exogeneity assumptions. Using the law of iterated expectations, we finally obtain (unconditional) unbiasedness $\mathrm{E}(\widehat{\boldsymbol{\gamma}}) = \boldsymbol{\gamma}$.

To implement the FE+ approach for the empirical example, the two steps are as follows: (1) estimate β_1 and β_2 by FE and (2) regress quasi-residuals, $r_{ij} \equiv y_{ij} - \widehat{\beta}_{1FE}x_{ij} - \widehat{\beta}_{2FE}w_jx_{ij}$, on w_j to obtain estimates of γ_0 and γ_1.

4.2 PC estimator

In this section, we define our proposed PC estimator. This estimator is best understood by using the two-stage formulation (see Section 2.1) of the general model in Equation 4, which requires some new notation.

The columns of the matrix $\mathbf{X}_j$ of unit-level covariates can be ordered, so that the matrix can be decomposed as $\mathbf{X}_j = (\mathbf{X}_{1j}\ \mathbf{X}_{2j}\ \mathbf{X}_{3j})$, where $\mathbf{X}_{1j}$ contains the R_1 unit-level covariates that have random slopes, so that $\mathbf{Z}_j = (\mathbf{1}_{n_j}\ \mathbf{X}_{1j})$, $\mathbf{X}_{2j}$ are R_2 cross-level interactions between unit-level covariates in $\mathbf{X}_{1j}$, and cluster-level covariates in $\mathbf{W}_j$, and $\mathbf{X}_{3j}$ are the R_3 remaining unit-level covariates (which may include cross-level interactions between covariates in $\mathbf{W}_j$ and other covariates in $\mathbf{X}_{3j}$). Correspondingly, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1',\ \boldsymbol{\beta}_2',\ \boldsymbol{\beta}_3')$ and $R = R_1 + R_2 + R_3$. The two-stage model can then be written as

$$\mathbf{y}_j = \mathbf{X}_{3j}\boldsymbol{\beta}_3 + \mathbf{Z}_j\boldsymbol{\eta}_j + \boldsymbol{\varepsilon}_j, \tag{15}$$

$$\eta_{rj} = \mathbf{w}_{rj}'\boldsymbol{\alpha}_r + u_{rj}, \quad r = 0, \ldots, R_1, \tag{16}$$

where $\boldsymbol{\eta}_j$ are cluster-specific coefficients. Here, Equation 15 is referred to as the Level-1 or unit-level model, and Equation 16 for the $R_1 + 1$ cluster-specific coefficients of the Level-1 model is referred to as the Level-2 or cluster-level model. When $r = 0$, Equation 16 is the model for the cluster-specific intercept η_{0j}, $\mathbf{w}_{0j}'$ is a row of $\mathbf{W}_j$ (with $\mathbf{W}_j = \mathbf{1}_{n_j} \otimes \mathbf{w}_{0j}'$) and $\boldsymbol{\alpha}_0 = \boldsymbol{\gamma}$. When $r > 0$, Equation 16 is the model for the cluster-specific slope of the rth column of $\mathbf{X}_{1j}$. The vector $\mathbf{w}_{rj}$ includes only those cluster-level covariates that interact with the rth column of $\mathbf{X}_{1j}$. The first element of $\boldsymbol{\alpha}_r$ is β_{1r} and the other elements are subsets of $\boldsymbol{\beta}_2$. Note that we include Level-2 models only for the coefficients of those unit-level covariates that have random slopes.

Equations 15 and 16 are a generalization of the two-stage formulation of the model for the HSB example in Equations 1 and 2. In that model, there are $R_1 = 1$ covariates $\mathbf{X}_{1j} = \mathbf{x}_j$ with random slopes and $\boldsymbol{\beta}_1 = \beta_1$. There are also $R_2 = 1$ cross-level interactions between unit-level covariates in $\mathbf{X}_{1j}$ and a cluster-level covariate w_j, namely $\mathbf{X}_{2j} = w_j\mathbf{x}_j$ with $\boldsymbol{\beta}_2 = \beta_2$. There are no remaining columns of $\mathbf{X}_j$, so $R_3 = 0$, and there is no $\mathbf{X}_{3j}$ or $\boldsymbol{\beta}_3$. Further, $\boldsymbol{\eta}_j = (\eta_{0j}, \eta_{1j})'$, $\mathbf{w}_{0j} = \mathbf{w}_{1j} = (1, w_j)'$, $\boldsymbol{\alpha}_0 = (\gamma_0, \gamma_1)'$, and $\boldsymbol{\alpha}_1 = (\beta_1, \beta_2)'$. We now outline the steps for our proposed PC estimator.

Step 1: Estimation of $\boldsymbol{\beta}_3$

Wooldridge (2005, 2010, section 11.7.2) considers the special case of this model without cluster-level covariates, that is, with $\boldsymbol{\eta}_j = \boldsymbol{\beta}_1 + \mathbf{u}_j$, and describes estimation of $\boldsymbol{\beta}_3$ (coefficients for all unit-level covariates without random slopes) by an extension of the de-meaning transformation used in FE estimation. Instead of premultiplying by the de-meaning operator $\mathbf{Q}_j$, we premultiply the Level-1 model by the projection matrix

$$\mathbf{M}_j \equiv \mathbf{I}_{n_j} - \mathbf{Z}_j\left(\mathbf{Z}_j'\mathbf{Z}_j\right)^{-1}\mathbf{Z}_j'.$$

Defining $\dot{\mathbf{y}}_j = \mathbf{M}_j\mathbf{y}_j$, $\dot{\mathbf{X}}_{3j} = \mathbf{M}_j\mathbf{X}_{3j}$, and $\dot{\boldsymbol{\varepsilon}}_j = \mathbf{M}_j\boldsymbol{\varepsilon}_j$, and noting that $\dot{\mathbf{Z}}_j = \mathbf{M}_j\mathbf{Z}_j = \mathbf{0}$, gives

$$\dot{\mathbf{y}}_j = \dot{\mathbf{X}}_{3j}\boldsymbol{\beta}_3 + \boldsymbol{\varepsilon}_j.$$

The POLS estimator of $\boldsymbol{\beta}_3$, denoted $\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}$, can be expressed as

$$\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}} = \boldsymbol{\beta}_3 + \left(J^{-1}\sum_{j=1}^{J}\dot{\mathbf{X}}_{3j}'\dot{\mathbf{X}}_{3j}\right)^{-1}\left(J^{-1}\sum_{j=1}^{J}\dot{\mathbf{X}}_{3j}'\dot{\boldsymbol{\varepsilon}}_j\right),$$

where $\sum_j \dot{\mathbf{X}}_{3j}'\dot{\mathbf{X}}_{3j}$ is assumed to be nonsingular with probability 1. If the ε_{ij} are normally distributed, this estimator also corresponds to the conditional ML (CML) estimator, conditioning on the sufficient statistics $\mathbf{Z}_j'\mathbf{y}_j$ for the "nuisance parameters" η_j (Verbeke, Spiessens, & Lesaffre, 2001).

The conditional expectation of $\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}$, given all covariates $\mathbf{V}$, becomes

$$\mathrm{E}\left(\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}|\mathbf{V}\right) = \boldsymbol{\beta}_3 + \left(J^{-1}\sum_{j=1}^{J}\dot{\mathbf{X}}_{3j}'\dot{\mathbf{X}}_{3j}\right)^{-1}\left(J^{-1}\sum_{j=1}^{J}\dot{\mathbf{X}}_{3j}'\mathrm{E}\left(\dot{\boldsymbol{\varepsilon}}_j|\mathbf{V}\right)\right).$$

Unit-level exogeneity, which implies that $\mathrm{E}\left(\dot{\boldsymbol{\varepsilon}}_j|\mathbf{V}\right) = \mathbf{0}$, is a sufficient condition for conditional unbiasedness $\mathrm{E}\left(\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}|\mathbf{V}\right) = \boldsymbol{\beta}_3$.

Step 2: Estimation of $\boldsymbol{\eta}_j$

Next, form quasi-residuals as

$$\mathbf{r}_j \equiv \mathbf{y}_j - \mathbf{X}_{3j}\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}$$

and then obtain OLS estimates $\breve{\boldsymbol{\eta}}_j$ for the regressions of $\mathbf{r}_j$ on $\mathbf{Z}_j$ for each cluster, $j = 1, \ldots, J$,

$$\breve{\boldsymbol{\eta}}_j = \left(\mathbf{Z}_j\mathbf{Z}_j'\right)^{-1}\mathbf{Z}_j'\mathbf{r}_j = \left(n_j^{-1}\sum_{i=1}^{n_j}\mathbf{z}_{ij}\mathbf{z}_{ij}'\right)^{-1}\left(n_j^{-1}\sum_{i=1}^{n_j}\mathbf{z}_{ij}r_{ij}\right), \tag{17}$$

where $\mathbf{z}'_{ij}$ is the ith row of $\mathbf{Z}_j$, and $\sum_{i=1}^{n_j} \mathbf{z}_{ij}\mathbf{z}'_{ij}$ is nonsingular with probability 1, which requires that $R_1 + 1 \leq n_j$. This step gives rise to the name PCs. Identical estimates of $\boldsymbol{\beta}_3$ and $\boldsymbol{\eta}_j$ are obtained by treating $\boldsymbol{\eta}_j$ as fixed parameters in Equation 15 via the inclusion of dummy variables for the clusters and interactions between these dummy variables and the columns of $\mathbf{Z}_j$.

The estimator in Equation 17 can alternatively be expressed as

$$\breve{\boldsymbol{\eta}}_j = \boldsymbol{\eta}_j + \left(\mathbf{Z}_j\mathbf{Z}'_j\right)^{-1}\mathbf{Z}'_j\left[\mathbf{X}_{3j}\left(\boldsymbol{\beta}_3 - \widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}\right) + \boldsymbol{\varepsilon}_j\right], \tag{18}$$

and the conditional expectation of $\breve{\boldsymbol{\eta}}_j$, given $\mathbf{V}$, becomes

$$\mathrm{E}\left(\breve{\boldsymbol{\eta}}_j|\mathbf{V}\right) = \boldsymbol{\eta}_j + \left(\mathbf{Z}_j\mathbf{Z}'_j\right)^{-1}\mathbf{Z}'_j\left\{\mathbf{X}_{3j}\left[\boldsymbol{\beta}_3 - \mathrm{E}\left(\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}|\mathbf{V}\right)\right] + \mathrm{E}\left(\boldsymbol{\varepsilon}_j|\mathbf{V}\right)\right\},$$

where $\mathbf{Z}_j\mathbf{Z}'_j$ is assumed to be nonsingular with probability 1. Because $\mathrm{E}\left(\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}|\mathbf{V}\right) = \boldsymbol{\beta}_3$ from Step 1, it follows that $\mathbf{X}_{3j}\left[\boldsymbol{\beta}_3 - \mathrm{E}\left(\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}|\mathbf{V}\right)\right] = \mathbf{0}$. It follows from unit-level exogeneity that $\mathrm{E}\left(\boldsymbol{\varepsilon}_j|\mathbf{V}\right) = \mathbf{0}$, and therefore $\breve{\boldsymbol{\eta}}_j$ is conditionally unbiased, $\mathrm{E}\left(\breve{\boldsymbol{\eta}}_j|\mathbf{V}\right) = \boldsymbol{\eta}_j$.

Step 3: Estimation of $\boldsymbol{\gamma}$, $\boldsymbol{\beta}_1$, and $\boldsymbol{\beta}_2$

The remaining regression coefficients $\boldsymbol{\gamma}$ (for cluster-level covariates), $\boldsymbol{\beta}_1$ (for unit-level covariates with random slopes), and $\boldsymbol{\beta}_2$ (for cross-level interactions involving unit-level covariates with random slopes) are now estimated. These coefficients are contained in the vectors $\boldsymbol{\alpha}_r$, $r = 0, \ldots, R_1$, in Equation 16. We write each Level-2 equation for all clusters using the following vector notation. Let $\boldsymbol{\eta}^*_r = (\eta_{r1}, \ldots, \eta_{rJ})'$ and $\mathbf{u}^*_r = (u_{r1}, \ldots, u_{rJ})'$ and let $\mathbf{W}^*_r$ have J rows $\mathbf{w}'_{rj}$, $j = 1, \ldots, J$. The Level-2 equation for each $\boldsymbol{\eta}^*_r$ can then be written as

$$\boldsymbol{\eta}^*_r = \mathbf{W}^*_r\boldsymbol{\alpha}_r + \mathbf{u}^*_r, \quad r = 0, \ldots, R_1.$$

Denoting the vector of estimates $\breve{\boldsymbol{\eta}}^*_r \equiv \left(\breve{\eta}_{r1}, \ldots, \breve{\eta}_{rJ}\right)'$, the model can be written as

$$\boldsymbol{\eta}^*_r = \mathbf{W}^*_r\boldsymbol{\alpha}_r + \mathbf{u}^*_r + \breve{\boldsymbol{\eta}}^*_r - \boldsymbol{\eta}^*_r$$

We estimate $\boldsymbol{\alpha}_r$ by applying OLS to the regression of $\boldsymbol{\eta}^*_r$ on $\mathbf{W}^*_r$, giving

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_r &= \boldsymbol{\alpha}_r + \left(\mathbf{W}^{*\prime}_r\mathbf{W}^*_r\right)^{-1}\mathbf{W}^{*\prime}_r\left(\mathbf{u}^*_r + \breve{\boldsymbol{\eta}}^*_r - \boldsymbol{\eta}^*_r\right) \\ &= \boldsymbol{\alpha}_r + \left(J^{-1}\sum_{j=1}^{J}\mathbf{w}_{rj}\mathbf{w}'_{rj}\right)^{-1}\left(J^{-1}\sum_{j=1}^{J}\mathbf{w}_{rj}\left(u_{rj} + \breve{\eta}_{rj} - \eta_{rj}\right)\right),\end{aligned} \tag{19}$$

where, for each r, $\sum_j \mathbf{w}_{rj}\mathbf{w}'_{rj}$ is assumed to be nonsingular with probability 1.

The conditional expectation of $\widehat{\boldsymbol{\alpha}}_r$, given $\mathbf{V}$, is given by:

TABLE 1
Estimates of Regression Coefficients From Different Methods For HSB Data

	Random Effects (RE)		Augmented Fixed Effects (FE+)		Per-Cluster Regression (PC)	
	Estimates	(*SE*)	Estimates	(*SE*)	Estimates	(*SE*)
γ_0 [Constant]	11.752	(0.232)	11.769	(0.205)	11.615	(0.271)
γ_1 [Catholic]	2.130	(0.346)	2.186	(0.337)	2.253	(0.406)
β_1 [SES]	2.958	(0.143)	2.782	(0.145)	2.772	(0.169)
β_2 [SES × Catholic]	−1.313	(0.216)	−1.349	(0.218)	−1.303	(0.234)

Note. All estimates are significantly different from 0 at the 0.05 level. HSB = high school and beyond.

$$\mathrm{E}(\widehat{\boldsymbol{\alpha}}_r|\mathbf{V}) = \boldsymbol{\alpha}_r + \left(J^{-1}\sum_{j=1}^{J}\mathbf{w}_{rj}\mathbf{w}_{rj}'\right)^{-1}\left(J^{-1}\sum_{j=1}^{J}\mathbf{w}_{rj}\left[\mathrm{E}\left(u_{rj}|\mathbf{V}\right)+\mathrm{E}\left(\breve{\eta}_{rj}|\mathbf{V}\right)-\eta_{rj}\right]\right).$$

It follows from cluster-level exogeneity that $\mathrm{E}\left(u_{rj}|\mathbf{V}\right)=0$ and from the results for Step 2 that $\mathrm{E}\left(\breve{\eta}_{rj}|\mathbf{V}\right)=\eta_{rj}$. Hence, $\mathrm{E}(\widehat{\boldsymbol{\alpha}}_r|\mathbf{V})=\boldsymbol{\alpha}_r$, and using the law of iterated expectations, we see that the estimator is unbiased; $\mathrm{E}(\widehat{\boldsymbol{\alpha}}_r)=\boldsymbol{\alpha}_r$.

For the special case of our model with $\boldsymbol{\eta}_j=\boldsymbol{\beta}_1+\mathbf{u}_j$, the estimator for $\boldsymbol{\beta}$ becomes the sample mean of $\breve{\boldsymbol{\eta}}_r^*$, and that estimator has been proposed by Wooldridge (2010, equation 11.80). In models in which $\mathbf{X}_j=\left(\mathbf{X}_{1j}\ \mathbf{X}_{2j}\right)$, or $R_3=0$, the first step can be skipped and $\mathbf{r}_j=\mathbf{y}_j$. Our empirical illustration is an example of the latter special case. Accordingly, we first estimate η_{0j} and η_{1j} in Equation 2 for each cluster j by regressing $\mathbf{y}_j$ on $\mathbf{x}_j$ using OLS, giving unbiased estimates $\breve{\eta}_{0j}$ and $\breve{\eta}_{1j}$. Identical estimates are obtained by OLS with dummy variables for clusters and interactions between these dummy variables and x_{ij}. Next, $\breve{\eta}_{0j}$ and $\breve{\eta}_{1j}$ are both regressed on w_j using OLS. In the regression for $\breve{\eta}_{0j}$, the OLS estimator for the intercept is unbiased for γ_0 and the OLS estimator for the coefficient of w_j is unbiased for γ_1. In the regression for $\breve{\eta}_{1j}$, the OLS estimator for the intercept is unbiased for β_1 and the OLS estimator for the coefficient of w_j is unbiased for β_2. If we did not include the cross-level interaction term, $x_{ij}w_j$, in our model, there would be no β_2, $R_1=R$, and we would regress $\breve{\eta}_{1j}$ on just the intercept, that is, find its sample mean, to obtain the unbiased estimate of β_1.

5 Empirical Example

To ground comparisons of our estimators of interest, we apply each to the HSB data introduced in Section 2.1. Table 1 provides estimates of the regression

coefficients for Equation 3. All estimates were obtained using standard commands in Stata 13 (StataCorp, 2013), such as mixed and xtreg (see Online Appendix B). Note that the RE estimate of the correlation between the random intercept and slope is 1, a relatively frequent occurrence in random-coefficient models (Chung, Gelman, Rabe-Hesketh, Liu, & Dorie, In press).

Castellano et al. (2014) show that positive correlation between a random intercept and a student-level covariate leads to overestimation of the coefficient of the covariate. Indeed, from the HSB data results presented in Table 1, we see that RE produces the largest estimate of the coefficient of SES, 2.958, approximately 6% higher than the closest estimate (FE+). The indicator variable w_j for Catholic schools is positively correlated with SES and therefore overestimation of the coefficient of SES is accompanied by underestimation of the coefficient of w_j, with RE producing the smallest estimate of γ_1, at 2.130.

While the differences in the FE+ and RE estimates of γ_1 may be practically significant, they are close in magnitude to the estimated standard errors (*SE*s) of the coefficient estimates. FE+ produces estimates of both β_1 and γ_1 that lie between the estimates produced by RE and PC, which is intuitive, given that FE+ relies only on the uncorrelated variance assumption, whereas RE additionally requires exogeneity, and PC requires neither assumption. PC gives the smallest estimated effect of SES on math achievement scores ($\widehat{\beta}_1 = 2.772$) and the largest estimated effect of Catholic schooling ($\widehat{\gamma}_1 = 2.253$). These estimates differ by about 6% from the RE counterparts, enough to give practitioners pause.

The small difference between estimates of β_1 from FE+ and PC suggests that the within-school variance in SES is not strongly correlated with the random slope. In fact, the within-school standard deviation of SES has a correlation of only 0.04 with the estimated residuals from the regression of $\breve{\eta}_{1j}$ on w_j in the final step of the PC approach.

6 Simulation Study

We now conduct a simulation study to investigate the performance of the RE, FE+, and PC estimators. In particular, we are interested in the amount of bias for RE and FE+ when the respective assumptions of cluster-level exogeneity and uncorrelated variance are violated. We also evaluate all three estimators, RE, FE+, and PC, by their root mean square errors (RMSEs) and consider performance of the estimated *SE*s. We use Stata 13 (StataCorp, 2013) throughout.

6.1 Data Generation Process

We generate the data using our model of interest in Equation 3. We first draw the school-level variables for each of $J = 100$ clusters. The random intercepts u_{0j} and random slopes u_{1j} are drawn from a bivariate normal distribution with zero means and variance–covariance matrix defined by variances $\psi_0 = 0.4^2$ and

$\psi_1 = 0.25^2$ and correlation $\rho = 0.5$, giving the covariance $\psi_{10} = 0.05$. We specify these variances to reflect those found in our empirical example. The exogenous school-level covariate w_j is drawn independently from a normal distribution with mean 1.7 and variance $\sigma_w^2 = 1$.

We then generate the student-level covariate x_{ij} as

$$x_{ij} = b_0 u_{0j} + b_1 u_{1j} + b_2 w_j + a e_{ij}, \quad e_{ij} \sim N(0, \sigma_j), \tag{20}$$

where

$$a = \sqrt{1 - \psi_0 b_0^2 - \psi_1^2 b_1^2 - \sigma_w^2 b_2^2 - 2 b_0 b_1 \psi_{10}}.$$

Here, $b_0 = 1.33$, $b_1 = 2.13$, and $b_2 = 0.20$, so that x_{ij} is positively correlated with the random intercept, random slope, and school-level covariate w_j. Finally, we generate y_{ij} according to Equation 3 with $\gamma_0 = 1$, $\gamma_1 = 3$, $\beta_1 = 1$, and $\beta_2 = 2$.

The key assumption under which we want to assess the performance of the competing estimators is that the sample within-cluster variance s_j^2 of x_{ij} is uncorrelated with the random slope u_{1j}. Thus, the population within-cluster standard deviation σ_j is of particular importance. Accordingly, the uncorrelated variance assumption factor in this simulation has two levels: when it holds, $\sigma_j = 1$, and when it is violated, $\sigma_j = \exp(u_{1j})$.

Although our empirical example involves schools that tend to have large numbers of students, both RE and FE are commonly used with classrooms serving as clusters. Furthermore, there are numerous relevant applications with longitudinal data where we often find even smaller cluster sizes. Thus, we also vary cluster size, primarily considering cluster sizes of 4 and 20. For simplicity, we set cluster sizes equal across clusters, $n_j = n$. We fully cross the cluster size and uncorrelated variance assumption factors, yielding four primary simulation conditions defined by (large/small n) $\times$ (uncorrelated variance assumption holds/violated). To further determine the effect of cluster size when the uncorrelated variance assumption is violated, we also consider a range of cluster sizes from 4 to 50: $n = 4, 8, 14, 20, 50$.

All conditions are replicated 500 times. Due to occasional lack of variation of x_{ij} within some small clusters, the PC approach fails for some replications. The lowest number of successful replications is 489, which occurs when the variance of x_{ij} is correlated with the random slopes, and we have only four observations in each cluster. For all simulation conditions with a cluster size of 20, all 500 replications are successful.

6.2 Results

We evaluate the performance of each of our three estimators (RE, FE+, and PC) of the fixed regression coefficients in our model of interest (Equation 3) across our four simulation conditions. The estimated bias and RMSE are given

TABLE 2
Comparing Methods for Estimating the Coefficients Using Simulated Data

		β_1 $[x_{ij}]$		β_2 $[x_{ij} \times w_j]$		γ_1 $[w_j]$	
Simulation Condition	Method	100× Bias	100× RMSE	100× Bias	100× RMSE	100× Bias	100× RMSE
Small clusters and uncorrelated variance	RE	16.6*	21.6	0.8	7.0	−4.5*	8.0
	FE+	0.6	16.2	0.0	8.2	−0.3	7.4
	PC	1.9	25.5	−0.6	13.3	−0.5	12.7
Small clusters and correlated variance	RE	21.3*	24.2	1.2	6.1	−5.3*	8.2
	FE+	11.7*	19.1	0.2	7.9	−2.4*	8.2
	PC	−1.8	26.7	0.7	13.4	0.2	12.2
Large clusters and uncorrelated variance	RE	6.2*	10.0	0.2	3.9	−1.7*	4.9
	FE+	−0.3	8.0	0.2	4.0	0.2	5.3
	PC	−0.2	8.0	0.1	4.1	0.0	5.2
Large clusters and correlated variance	RE	12.6*	14.4	−0.1	3.5	−2.9*	5.2
	FE+	12.8*	15.2	−0.3	4.2	−2.5*	5.2
	PC	0.7	7.7	−0.3	4.0	0.0	5.0

Note. RMSE = root mean square error; RE = random effects; FE+ = augmented fixed effects; PC = per-cluster regression; small clusters: $n_j = n = 4$, large clusters: $n_j = n = 20$; uncorrelated variance: $\sigma_j^2 = 1$, correlated variance: $\sigma_j = \exp(u_{1j})$.
*Estimated bias differs significantly from 0 at the 0.05 level.

in Table 2. Online Appendix C provides supplemental tables for each coefficient that also include the mean *SE*s, standard deviations of the estimates, and the ratios of these values.

6.2.1 Bias. For β_1, the coefficient of the endogenous student-level covariate x_{ij}, there are three main results. First, the PC estimator is unbiased across all conditions even when the uncorrelated variance assumption is violated. Figure 1 clearly illustrates this finding as the empirical distributions of the errors (i.e., estimate − parameter) of the PC estimator (the solid curves) are centered on 0 in all four panels, where each panel represents one of the four simulation conditions.

Second, the RE estimator is biased regardless of whether the uncorrelated variance assumption holds, whereas the FE+ estimator is biased only when this assumption is violated. This result for RE is expected, given that the RE estimator relies on the assumption of both unit- and cluster-level exogeneity (see Section 3.2), and cluster-level exogeneity is violated in all four conditions, with the nonzero correlation between x_{ij} and both the random intercept and its random slope. We do note, however, that violation of the uncorrelated variance assumption exacerbates the magnitude of the RE estimator's bias: For the small cluster size condition ($n = 4$), the estimated bias

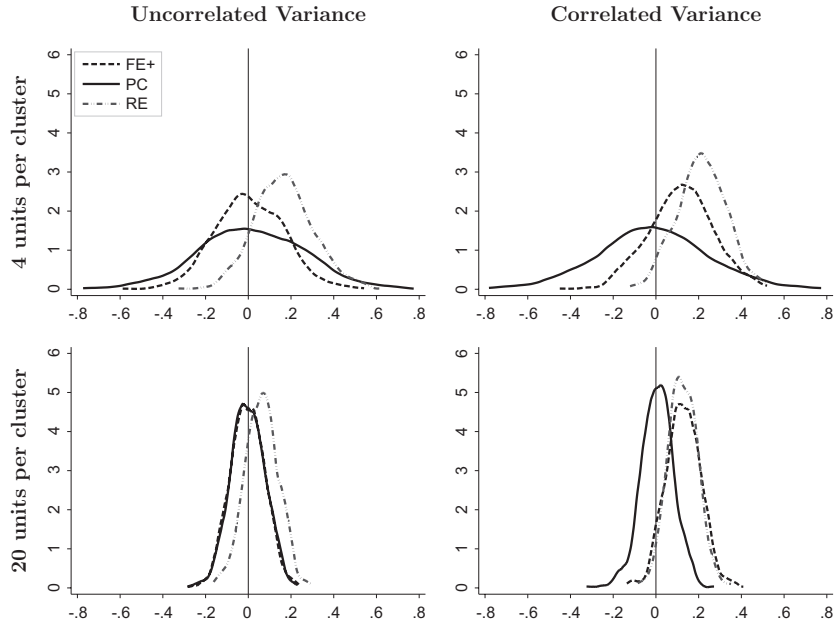

FIGURE 1. *Kernel density plots of estimation errors,* $\widehat{\beta}_1 - \beta_1$*, for coefficient of* x_{ij} *across replications for all methods when the uncorrelated variance assumption holds (left panels) and when it is violated (right panels).* Note. *FE+ = augmented fixed effects; PC = per-cluster regression; RE = random effects.*

is 1.28 times as large; and for the larger cluster size ($n = 20$), the estimated bias more than doubles as shown in the first column of results in Table 2. In contrast, the FE+ approach only requires unit-level exogeneity assumptions and thus produces unbiased estimates under cluster-level endogeneity as long as there is no correlation between the random slopes and within-cluster variance of x_{ij}. This is evident in Figure 1 by observing that the curves for FE+ (dashed) are more similar to those for PC (solid) in the left-hand plots (for uncorrelated variance simulation conditions) and more similar to the curves for RE (dot-dashed) in the right-hand plots (for correlated variance simulation conditions).

Third, the estimated bias for β_1 is larger than that for the other two regression coefficients, which is not surprising, given that x_{ij} is the source of the endogeneity. For instance, as shown in Table 2, the estimated bias of $\widehat{\beta}_{1\mathrm{RE}}$ ranges from 6.2% to 21.3% of the true value. The next largest estimated bias is -0.053 for $\hat{\gamma}_{1\mathrm{RE}}$ under the small clusters and uncorrelated variance condition, which is only 1.8% of the coefficient's true value ($\gamma_1 = 3$).

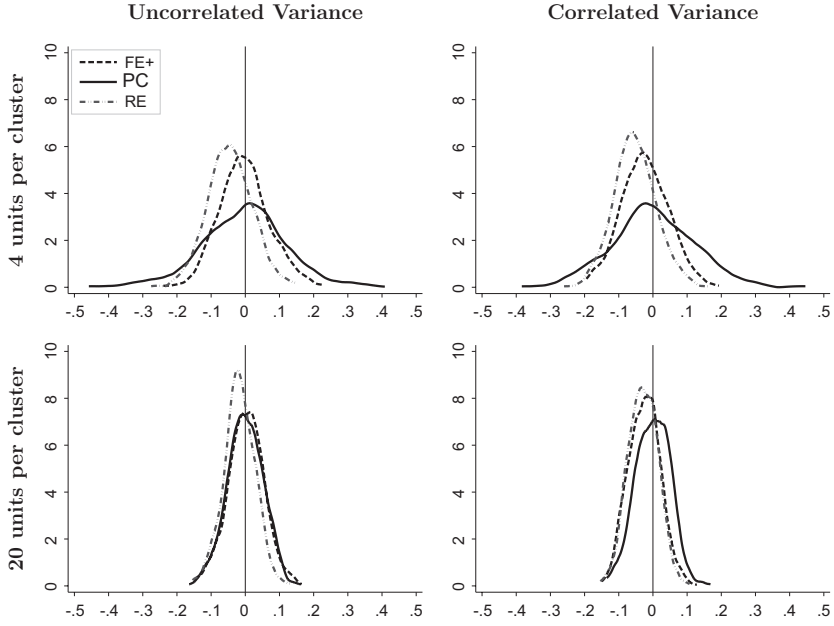

FIGURE 2. *Kernel density plots of estimation errors,* $\hat{\gamma} - \gamma_1$*, for coefficient of* w_j *across replications for all methods when the uncorrelated variance assumption holds (left panels) and when it is violated (right panels).* Note. *FE+ = augmented fixed-effects; PC = per-cluster regression; RE = random effects.*

The coefficient of the interaction term, β_2, is the least affected by the simulation conditions. We only find statistically significant bias (at the 5% level) for $\widehat{\beta}_{2RE}$ for the small cluster size condition—both when the uncorrelated variance assumption holds and when it is violated. Even in these cases, as given in Table 2, the estimated bias is rather small relative to the magnitude of the true value ($\beta_2 = 2$): It is 0.4% of the parameter value when the condition holds and 0.6% when it is violated. (Plots of the empirical distributions of the estimation errors for β_2 are given in Figure C.1 in Online Appendix C.)

The estimated biases of the estimators for the coefficient γ_1 of the exogenous school-level covariate w_j follow similar patterns as for the coefficient β_1 of the endogenous student-level covariate x_{ij}. Just as for β_1, the PC estimator is unbiased across all conditions, the FE+ estimator is biased only when the variance of x_{ij} is correlated with the random slope u_{1j} (i.e., uncorrelated variance assumption violated), and the RE estimator is biased regardless of whether the uncorrelated variance assumption is violated. These findings are clearly illustrated in Figure 2 by comparing the centers of the empirical distributions of errors for all estimators across all conditions: the PC curve

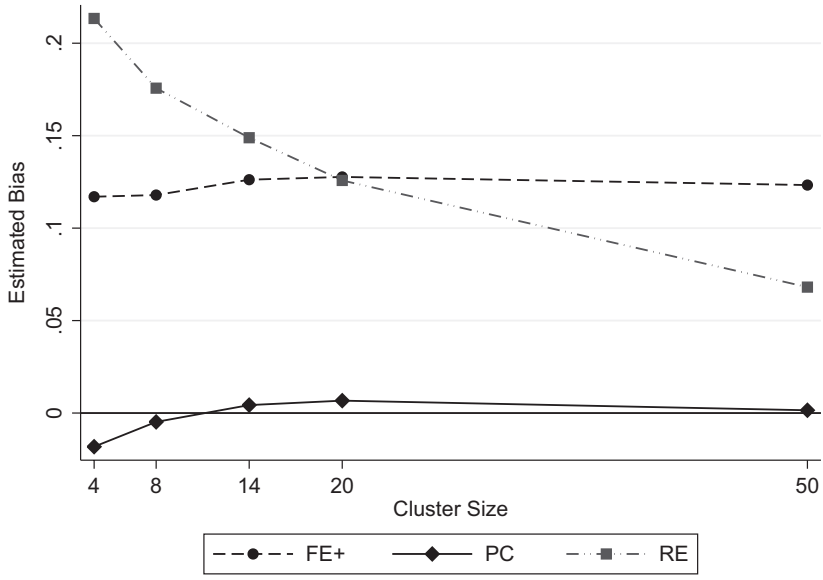

FIGURE 3. *Estimated bias for coefficient* β_1 *of* x_{ij} *versus cluster size.* Note. *FE+ = augmented fixed effects; PC = per-cluster regression; RE = random effects.*

(solid) is always centered on 0, whereas the RE curve (dot-dashed) is always centered below 0, and the FE+ curve (dashed) is centered below 0 only for the correlated variance conditions in the right-hand panels. Just as with β_1, the FE+ estimator's bias for γ_1 does not vary with cluster size—its estimate is about 0.8% of the true parameter value for both $n = 4$ and $n = 20$ as seen in Table 2. Cluster size affects the RE estimator's bias for γ_1, as it did for β_1: As cluster size increases, the bias decreases. When the uncorrelated variance assumption holds, this bias decreases by about 63% going from $n = 4$ to $n = 20$, and by about 45% when the assumption is violated (see Table 2).

Given that β_1 was most affected by the violation of the uncorrelated variance assumption, we further investigated the effect of cluster size on the estimates of this regression coefficient. Figure 3 gives the estimated bias for each estimator across cluster sizes of 4, 8, 14, 20, and 50. The PC (solid) curve hugs the $y = 0$ line. The FE+ and RE curves cross at $n = 20$: As cluster size increases, the RE estimator's bias decreases (dot-dashed curve), whereas the FE+ estimator's bias is not as affected by cluster size, shown by its dashed curve staying relatively constant across the range of cluster sizes. Thus, cluster size has a differential effect on the bias of the estimators. When using bias to evaluate the estimators, our simulation study provides strong evidence that our proposed PC estimator outperforms the other estimators.

6.2.2 Precision and RMSE. As is often the case, there is a trade-off between bias and precision, which depends in part on the size of the clusters. The rank ordering of the estimators by their standard deviation (*SD*) is approximately the same for the three regression coefficients with slight differences between the smaller and larger cluster size conditions. Accordingly, we discuss how the precision of the estimators depends on cluster size, without distinguishing among the coefficients.

For the smaller cluster sizes of $n = 4$, RE produces the estimates with the smallest variances, followed by FE+, and PC produces the most variable estimates. This is clearly illustrated by comparing the widths of the empirical distributions of errors in Figure 1 or 2 for each estimator: The RE curves are the narrowest and the PC curves are the widest. For instance, for $n = 4$ and when the uncorrelated variance assumption is violated, the *SD* of $\widehat{\beta}_{1PC}$ over 500 replications is about 0.266, whereas for RE, the *SD* is less than half that at about 0.114 (see Tables C.1–C.3 in Online Appendix C for all *SD* values).

For the larger cluster size of $n = 20$, RE always yields the smallest variances, but the variances are not much smaller than those for FE+ and PC, which tend to have about equal variances. For instance, for β_1, for the large clusters and uncorrelated variance condition shown in the lower, left-hand panel of Figure 1, it is difficult to discern any differences in the widths of the distributions. Indeed, the *SD* for RE, in this case, is about 0.078 and the *SD*s for both FE+ and PC are 0.080.

With regard to precision, RE consistently outperforms FE+ and PC for all the regression coefficients and across all the simulation conditions. However, given the trade-off between bias and precision, it is useful to evaluate the estimators with regard to their RMSEs, which take both bias and imprecision into account. Given that the estimates of β_1 are the most affected by the simulation conditions and that precision depends on cluster size, we consider the RMSEs as a function of the extended range of cluster sizes for β_1 in Figure 4. Just as with bias in Figure 3, Figure 4 shows that the FE+ and RE curves cross with RE outperforming FE+ as cluster size increases. This figure also shows that, for the smallest cluster size of 4, the RMSE for PC is large and similar to that of RE. However, with clusters of at least 8, the PC estimator outperforms both RE and FE+ with regard to RMSE, providing strong evidence in favor of the PC estimator.

6.2.3 SE Estimation. As a final point, we evaluate the estimators in terms of how well their estimated *SE*s approximate the sampling *SD*s. We again focus on the most affected regression coefficient β_1. Figure 5 displays this ratio of mean *SE* to *SD* over the extended range of cluster sizes—similar to Figures 3 and 4. If the *SE* estimation works well, this ratio should equal 1. We see that both the PC (solid curve) and RE (dot-dashed curve) approaches provide good *SE* estimates. In contrast, for the FE+ approach, the *SE*s are increasingly underestimated as the

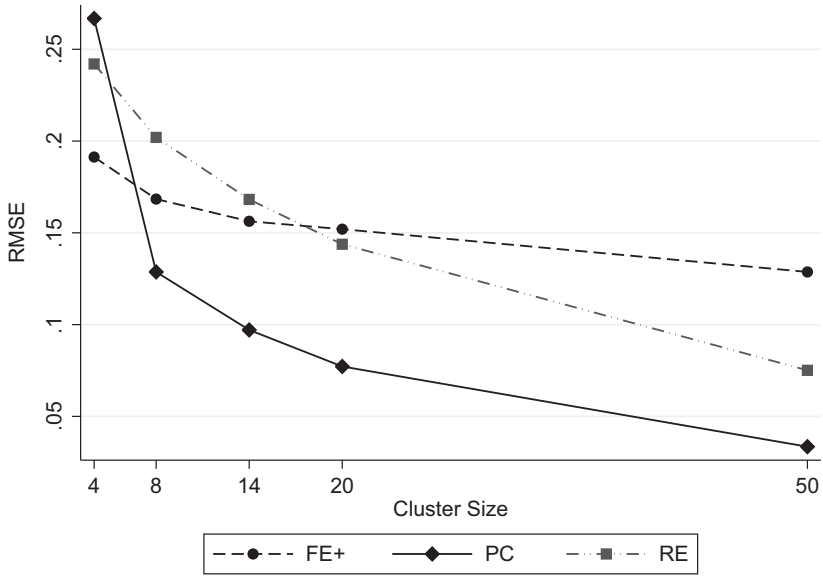

FIGURE 4. *Estimated root mean square error (RMSE) for coefficient* β_1 *of* x_{ij} *versus cluster size.* Note. *FE+ = augmented fixed effects; PC = per-cluster regression; RE = random effects.*

cluster size increases. Although both the FE+ and PC approaches treat estimated coefficients from previous steps as known in the subsequent steps, it appears that underestimation of the *SE* is a larger problem for the FE+ approach. Accordingly, we recommend using either analytically derived or bootstrapped *SE*s for the FE+ approach. These could also be used for the PC approach and may be necessary if Step 1 is required.

7 Conclusion

Given the popularity of multilevel models, studies that investigate potential biases for key parameters and provide simple solutions are clearly important. We have shown that commonly used random effects and FE estimators are biased in the presence of correlation between random effects and the within-cluster variance of unit-level covariates. Further, such bias can spill over to the estimation of coefficients of other covariates. We have proposed a new PC estimator that avoids such bias, produces good estimates of *SE*s, and generally has low RMSE. Consequently, we recommend broad use of PC when working with longitudinal or nested cross-sectional data when the clusters are sufficiently large. Stata code for applying this method to the HSB data is provided in Online Appendix B. In instances where the cluster sizes are small

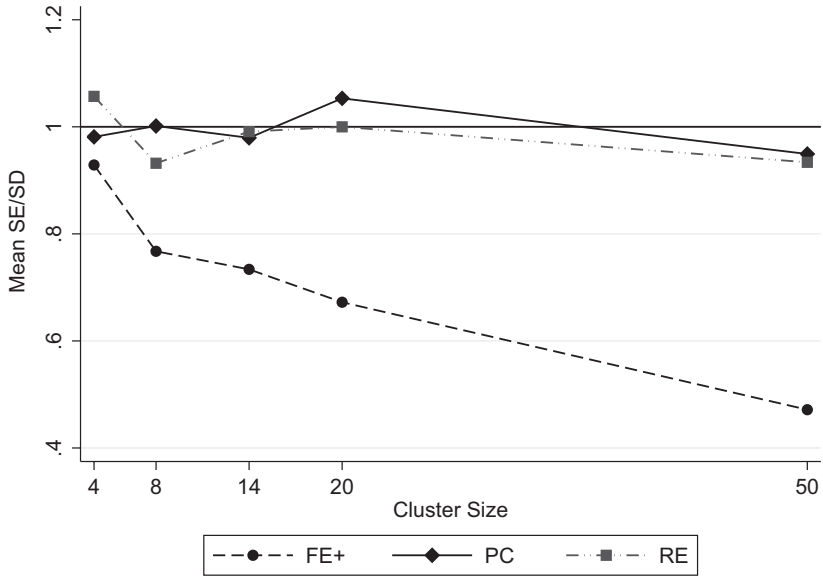

FIGURE 5. *Ratio of mean standard errors (mean* SE*s) divided by standard deviations* (SD*s*) *of estimates versus cluster size.* Note. *FE+ = augmented fixed effects; PC = per-cluster regression; RE = random effects.*

relative to the number of random effects, or where estimates for the random part of the model are of interest, we recommend using PC as part of a sensitivity analysis for alternative estimators.

Per-cluster methods have been used in the past for linear multilevel models (Burstein, Linn, & Capell, 1978, p. 369) and multilevel structural equation models (Chou, Bentler, & Pentz, 2000). Per-cluster methods can also be used for nonlinear multilevel models, such as probit models with random intercepts (Borjas & Sueyoshi, 1994) and logit models with random intercepts and slopes (Korn & Whittemore, 1979). However, the purpose of that work was to develop simple estimators and not to address endogeneity concerns. For our proposed PC estimator for linear models, it might appear to be inefficient to use OLS in the final step, not taking into account that the intercepts and slopes are estimated with different precision for different clusters. However, FGLS approaches, such as those discussed by Berkey, Hoaglin, Antczak-Bouckoms, Mosteller, and Golditz (1998), suffer from similar biases as RE estimators, as we confirmed in simulations (not shown).

An alternative approach for handling endogeneity, proposed for random-intercept models by Allison and Bollen (1997) and Teachman, Duncan, Yeung, and Levy (2001), is to model the unit-level covariates jointly with the responses

using structural equation modeling and allow them to be correlated with the random intercept. This approach can be generalized to random-coefficient models but becomes infeasible for large cluster sizes.

In summary, we have demonstrated that our proposed, simple-to-implement PC approach outperforms standard estimators when estimating regression coefficients in multilevel models under violations of both the cluster-level exogeneity and uncorrelated variance assumptions. We recommend that researchers consider the validity of the uncorrelated variance assumption and add the PC method to their toolbox when investigating effects of covariates in cross-sectional and longitudinal analyses.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grants R305B090011 to Michigan State University and R305B110017 to University of California, Berkeley. The opinions expressed are those of the authors and do not represent the views of the Institute or of the U.S. Department of Education.

Supplemental Material

The online appendices are available at http://jeb.sagepub.com/supplemental.

References

Allison, P. D., & Bollen, K. A. (1997). *Change score, fixed effects, and random component models: A structural equation approach*. Paper presented at the Annual Meeting of the American Sociological Association.

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background variables in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, *29*, 37–65.

Berkey, C. S., Hoaglin, D. C., Antczak-Bouckoms, A., Mosteller, F., & Golditz, G. A. (1998). Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine*, *17*, 2537–2550.

Borjas, G. J., & Sueyoshi, G. T. (1994). A two-stage estimator for probit models with structural group effects. *Journal of Econometrics*, *64*, 165–182.

Burstein, L., Linn, R. L., & Capell, F. J. (1978). Analyzing multilevel data in the presence of heterogeneous within-class regressions. *Journal of Educational Statistics*, *3*, 347–383.

Castellano, K. E., Rabe-Hesketh, S., & Skrondal, A. (2014). Composition, context, and endogeneity in school and teacher comparisons. *Journal of Educational and Behavioral Statistics*, *39*, 333–367.

Chou, C.-P., Bentler, P. M., & Pentz, M. A. (2000). Two-stage approach to multilevel structural equation models: Application to longitudinal data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 33–49). Mahwah, NJ: Lawrence Erlbaum.

Chung, Y., Gelman, A., Rabe-Hesketh, S., Liu, J., & Dorie, V. (In press). Weakly informative prior for point estimation of covariance matrices in hierarchical models. *Journal of Educational and Behavioral Statistics*.

Don, F., & Magnus, J. (1980). On the unbiasedness of iterated GLS estimators. *Communications in Statistics—Theory and Methods*, *9*, 519–527.

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

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

Kakwani, N. C. (1967). The unbiasedness of Zellner's seemingly unrelated regression equations estimators. *Journal of the American Statistical Association*, *62*, 141–142.

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

Korn, E. L., & Whittemore, A. S. (1979). Methods for analyzing panel studies of acute health effects of air pollution. *Biometrics*, *35*, 795–802.

Mundlak, Y. (1978). On the pooling of time series and cross-sectional data. *Econometrica*, *46*, 69–86.

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

Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, *20*, 307–335.

Spanos, A. (2006). Where do statistical models come from? Revisiting the problem of specification. *IMS Lecture Notes—Monograph Series 2nd Lehmann Symposium—Optimality*, *49*, 98–119.

StataCorp. (2013). *Stata: Statistics/data analysis [computer software]*. College Station, TX: Author.

Teachman, J., Duncan, G. J., Yeung, W. J., & Levy, D. (2001). Covariance structure models for fixed and random effects. *Sociological Methods & Research*, *30*, 271–288.

Verbeke, G., Spiessens, B., & Lesaffre, E. (2001). Conditional linear mixed models. *American Statistician*, *55*, 25–34.

Wiley, D. E. (1975). Another hour, another day: Quantity for schooling, a potent path for policy. In R. M. Hauser, W. H. Sewell, & D. Alwin (Eds.), *Schooling and achievement in American Society* (pp. 225–265). New York, NY: Academic Press.

Wooldridge, J. M. (2005). Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. *Review of Economics and Statistics*, *51*, 385–390.

Wooldridge, J. M. (2010). *Econometric analysis of cross-section and panel data*. Cambridge, MA: The MIT Press.

AUTHORS

MICHAEL DAVID BATES is a doctoral student in the Department of Economics at Michigan State University, 486 W. Circle Dr., Michigan State University, East Lansing, MI 48824; e-mail: batesmi3@msu.edu. His research interests include labor and public economics, empirical industrial organization, economics of education, and longitudinal data analysis in linear and nonlinear settings.

KATHERINE E. CASTELLANO is an associate psychometrician at Educational Testing Service, 90 New Montgomery St, Suite 1500, San Francisco, CA 94105; e-mail: kecastellano@ets.org. Her research interests include multilevel modeling and student growth models.

SOPHIA RABE-HESKETH is a professor at the Graduate School of Education and Graduate Group in Biostatistics, University of California, Berkeley, 3659 Tolman Hall, University of California, Berkeley, CA 94720; e-mail: sophiarh@berkeley.edu. Her primary research interests are multilevel and latent variable modeling.

ANDERS SKRONDAL is a senior scientist at the Division of Epidemiology, Norwegian Institute of Public Health, P.O. Box 4404 Nydalen, N-0403, Oslo, Norway; e-mail: anders.skrondal@fhi.no. His research interests include topics in statistics, biostatistics, social statistics, econometrics, and psychometrics including latent variable, multilevel, and longitudinal modeling.

Manuscript received May 31, 2014
Revision received September 30, 2014
Accepted October 10, 2014