# ADAPTIVE CENTERING WITH RANDOM EFFECTS: AN ALTERNATIVE TO THE FIXED EFFECTS MODEL FOR STUDYING TIME-VARYING TREATMENTS IN SCHOOL SETTINGS

**Stephen W. Raudenbush**
Department of Sociology
University of Chicago
1126 East 59th St., Room SS 416
Chicago, IL 60637
sraudenb@uchicago.edu

**Abstract**

Fixed effects models are often useful in longitudinal studies when the goal is to assess the impact of teacher or school characteristics on student learning. In this article, I introduce an alternative procedure: adaptive centering with random effects. I show that this procedure can replicate the fixed effects analysis while offering several comparative advantages: the incorporation into standard errors of multiple levels of clustering; the modeling of heterogeneity of treatment effects; the estimation of effects of treatments at multiple levels; and computational simplicity. After illustrating these ideas in a simple setting, the article formulates a general linear model with adaptive centering and random effects and derives efficient estimates and standard errors. The results apply to studies that have an arbitrary number of nested and cross-classified factors such as time, students, classrooms, schools, districts, or states.

## 1. INTRODUCTION

This article revisits a familiar problem: the use of fixed effects models to facilitate causal inferences in longitudinal research. Motivating this reexamination is a series of important policy studies in education. Such studies might aim to assess the impact on student learning of certain teacher characteristics or instructional practices, the impact on students of attending private schools or charter schools, and the impact of new regulations governing school organization, resources, or incentives.

Educational policy studies of this kind commonly have a complex sample design. Repeatedly observed students will typically be moving across classrooms and teachers within a school over time, and they will often be migrating across schools or even districts over time as well. The interventions of interest are often defined at the teacher, school, or district level. The question then arises as to whether the fixed effects specification as described in classic econometric texts is the optimal specification for statistical inference. My conclusion is that we can realize the benefits of the fixed effects specification while overcoming some of its limitations by using an approach I call "adaptive centering with random effects." A brief review of relevant statistical history and current educational applications sets the stage for the argument.

### Statistical History

In foundational papers on the advantages of fixed effects analysis in longitudinal data (Hausman 1978; Mundlak 1978), the design of interest involves the collection of up to $T$ longitudinal measurements on each of $n$ units. These units might be persons, firms, states, or countries. During the time period of interest, some or all of these units are subject to an intervention or "treatment," and the aim is to assess the impact of this treatment on the continuous outcome $y_{ti}$ observed on unit $i$ at time $t$ for $i \in (1, \ldots, n); t \in (1, \ldots, T_i)$. Some authors have referred to such a design as an "interrupted time series" (Campbell and Stanley 1963). Its beauty derives from the fact that the same unit is observed under two or more treatment conditions; thus unobserved time-invariant characteristics of the units can be removed from the estimated treatment effect.

#### *The Fixed Effects Model*

A simple model for such data is

$$y_{ti} = \beta x_{ti} + u_i + e_{ti} \tag{1}$$

where $x_{ti}$ is the treatment status or "dosage," $\beta$ captures the association between the treatment and the outcome, $u_i$ is a fixed effect that represents the combined effect of all time-invariant influences on the outcome, and $e_{ti}$

is a zero-mean stochastic error term that varies over time within units but not between them. The fixed effects approach absorbs the impact of the $u_i$ by including dummy variables for each unit. It is therefore not necessary to assume away the influence of time-invariant, unit-specific confounding variables. Two key assumptions then identify $\beta$ as a causal effect: a functional form assumption requires that the causal effect of interest is linear in $x_{ti}$ and an assumption of no time-varying confounding, that is, $E(\varepsilon_{ti} \mid x_{ti}) = E(e_{ti}) = 0$. The assumption of no time-varying confounding can be relaxed by including measured time-varying covariates, though problems arise if such covariates are both outcomes of previous treatments and predictors of exposure to later treatments. Robins (2000) has developed methods for handling such time-varying confounders; Hong and Raudenbush (2008) have extended these methods to multilevel data. I do not consider time-varying confounders in detail in the current article.

#### *The Random Effects Model*

The standard alternative model is the random effects model, which adds an intercept $\theta$ to equation 1 and then regards the unit-specific effect $u_i$ as randomly varying across a population of units:

$$y_{ti} = \theta + \beta x_{ti} + u_i + e_{ti}, \tag{2}$$

where $u_i$ is now a zero-mean random variable. As Hausman (1978) noted, one advantage of the random effects approach is that it can make use of more information than can the fixed effects approach in estimating the treatment effect. Suppose, for example, that some units were randomly assigned to receive the same dosage of the treatment at every time point. The fixed effects approach would effectively drop such units from the analysis; the impact of the treatment on those units would be absorbed into the fixed effects $u_i$. In contrast, the random effects approach would exploit this information about the treatment effect. The random effects approach is also more flexible than the fixed approach in allowing the analyst to simultaneously study time-invariant interventions in terms of main effects and interactions with $x_{ti}$. Finally, it is straightforward to incorporate heterogeneity of treatment effects using random coefficients within the random effects approach (see Raudenbush and Bryk 2002 for a review).

However, these advantages of the random effects approach come at a price. Under the random effects model, one must assume that mean of $u_i$ does not depend upon $x_{ti}$: $E(u_i \mid x_{ti}) = E(u_i) = 0$. Therefore if any characteristic of unit $i$ that is associated with the outcome also predicts access to the treatment, that characteristic must be observed and included in equation 2. Effectively, one must assume that all unobserved confounding variables are independent of

treatment status once the observed confounders are controlled. The beauty of the fixed effects specification 1 is that it does not require this strong assumption: all time-invariant influences, including those not observable, are removed from the error.

A question arising naturally from this literature is whether it is possible to enjoy the best of both worlds: can we reap the key benefit of the fixed effects approach in eliminating time-invariant confounding while also incorporating some of the benefits of the random effects approach? We won't be able to buy the efficiency of the random effects approach without adding the key assumption of no unobserved time-invariant confounding. But perhaps we can incorporate the other benefits: multilevel treatments and treatment effect heterogeneity. This reasoning lays part of the basis for using adaptive centering with random effects. There is, however, an additional rationale: assuring that the error structure of the model correctly reflects the clustered nature of educational data.

### Recent Educational Applications

During the past decade, there has been an explosion of interest in using longitudinal test score data to study the effects of educational resources, practices, and policies. In large part, this exciting new genre of research is now possible because more and more states and school districts, responding to demands for accountability, have assembled sophisticated longitudinal data systems that annually test students and track them as they move across classrooms and schools. For example, Clotfelter, Ladd, and Vigdor (2007a, 2007b) used longitudinal data from North Carolina to study the impact of teacher qualifications and the distribution of teacher qualifications by race. Bifulco and Ladd (2006) used North Carolina data to study the impact of attending a charter school. Chicago's longitudinal data system has supported studies of remedial education policies (Jacob and Lefgren 2004a), teacher training (Jacob and Lefgren 2004b), high-stakes testing in elementary schools (Neal and Schanzenbach, in press) and in high schools (Roderick and Nagoka 2005). Harris and Sass (2008) exploited longitudinal data from Florida to study teacher certification. Hanushek, Kain, and Rivkin (2004) used Texas data to study student mobility and a variety of other policies (see review by Hanushek and Rivkin 2006).

Applications of this type typically involve a nested structure that is not explicitly represented in the statistical model. For example, in studies of the impact of qualifications of elementary school teachers, the treatment variable $x_{ti}$ will take on the same value for every student who shares membership in the same classroom in a given year. In a study of the impact of charter schools, the identification strategy is typically to use school fixed effects in order to isolate the within-student impact of moving from a charter school to

a non-charter school or back again. However, the students attending a given school at a given time create a cluster that is not represented in the model. In a time series of repeated cross sections of students, a collection of students will typically be clustered in classrooms within schools at a given time; the aim is to study changes in a school mean (or adjusted mean) over time as a function of changes in school policy or practice (Paterson 1991). Such a design has four levels (students, classrooms, schools, time) that may or may not be reflected in the model.

It is common to use Huber-White corrected standard errors to reflect the statistical dependence that arises from aspects of clustering that are not represented in the model (White 1980). However, this is often difficult. For example, in a study of students moving across schools, any student can, in principle, attend a school attended by another student who previously attended any other school. The result is that the vector of residuals for any student is potentially correlated with the vector of residuals of every other student, a fact that makes computation of the Huber-White standard errors daunting given the massive sample sizes of these studies.

A final characteristic of recent longitudinal policy studies in education is a perceived need to control for two or more dimensions of fixed effects. For example, in a study of the impact of teacher characteristics, one might want to control for school fixed effects as well as child fixed effects. The same is true in a study of students moving across schools, for example, a study of the impact of charter schools. When a study involves hundreds of thousands of students and hundreds or even thousands of schools, the computational complexity may become burdensome.

It would be ideal if one could devise a general analytic approach that, like the fixed effects approach, can absorb the contribution of unobserved, time-invariant confounding while also explicitly reflecting the clustered design of the sample and enabling efficient computation. Also, building on the historical discussion of the previous section, it would be good if certain benefits of the random effects model could be combined with the known benefits of the fixed effects model. In particular, it would be good to allow treatments (or explanatory variables) at multiple levels and to have a natural way to incorporate treatment effect heterogeneity.

### Claims

The article makes several claims:

1. By adaptively centering the treatment variable within the framework of a random effects model, it is possible to replicate the results of the fixed effects analysis. The replication is exact in the case of balanced data.

When the data are unbalanced, point estimates of treatment effects are identical to those of the fixed effects approach, while standard errors are slightly different in small samples while converging with the sample size.

2. Adaptive centering with random effects can naturally incorporate multiple sources of clustering that arise in educational data. By better reflecting the data collection design, the aim here is to obtain more efficient inference and more realistic standard errors.
3. The approach naturally extends to settings in which there are multiple treatments of interest and where these treatments vary at different levels. For example, one might be interested in whether the benefits of teacher knowledge are the same in public and private schools in a setting where school sector (public versus private) is time invariant.
4. The approach naturally extends to incorporate heterogeneity of treatment effects using random coefficients.
5. The computations are relatively straightforward.

### Organization

Section 2 of this article provides a heuristic motivation for the approach based on relevant literature, while section 3 illustrates how the approach works in the context of a simple hypothetical data set with known parameters. Section 4 then proposes a general linear model with an arbitrary number of levels of clustering. That section presents efficient estimators and standard errors within the framework of maximum likelihood. It then illustrates how the approach works with one dimension of confounding (e.g., time-invariant student-level confounders) and two dimensions of confounding (e.g., time-invariant student and school confounding). Extensions to an arbitrary number of dimensions of confounding follow straightforwardly.

## 2. HEURISTIC MOTIVATION

In this section, I show how adaptive centering with random effects is linked to familiar ideas in fixed effects, conventional ordinary least squares (OLS) regression, and hierarchical linear models. In general, the aim is to decompose a statistical association into its within- and between-context components. Allison (2005) provides a more extensive discussion.

### Centering in the Fixed Effects Model

One of the difficulties in estimating the parameters of the fixed effects model (equation 1) arises when there are many persons and few time points per person. That is, $n$ is large but $T$ is small. It would be awkward or even impossible to include a dummy variable for each person because such an approach might

involve thousands or even hundreds of thousands of parameters. Therefore, statistical packages exploit centering to perform the computations. We note that the person-specific mean of equation 1 is

$$\bar{y}_i = \beta \bar{x}_i + u_i + \bar{e}_i, \tag{3}$$

where $\bar{y}_i, \bar{x}_i, \bar{e}_i$ are the person-specific means of $y_{ti}, x_{ti}, e_{ti}$, respectively. Subtracting equation 3 from equation 1 yields

$$y_{ti} - \bar{y}_i = \beta(x_{ti} - \bar{x}_i) + e_{ti} - \bar{e}_i, \tag{4}$$

revealing that the fixed effects estimate of $\beta$ can be obtained by OLS where the centered outcome is regressed on the centered predictor. Standard errors and significance tests must be corrected for the loss of degrees of freedom (centering of the outcome induces correlations within persons among the errors $e_{ti} - \bar{e}_i$), but such corrections are trivial computationally.

#### Decompositions of Associations within and between Levels

The coefficient $\beta$ represents the within-person association between $x$ and $y$. The between-person association can be obtained from the regression of "means on means":

$$\bar{y}_i = \alpha + \beta_b \bar{x}_i + \bar{e}_i. \tag{5}$$

Here $\beta_b$ represents the between-person coefficient. Note that if $\beta_b = \beta$—that is, if the within and between coefficients are equal—the random effects model (equation 2) is justified: the between-person and the within-person information can be exploited to generate a more efficient estimate of $\beta$ than is possible using the within-person information alone, as in the case of the fixed effects model. Therefore a specification test for the appropriateness of the random effects model would be a test of $H_0 : \beta_b = \beta$.

Substituting equation 5 back into equation 4 then yields the contextual effects model (Firebaugh 1978; see Willms's 1986 review of applications):

$$y_{ti} = \alpha + \beta_b \bar{x}_i + \beta(x_{ti} - \bar{x}_i) + e_{ti} = \alpha + (\beta_b - \beta)\bar{x}_i + \beta x_{ti} + e_{ti}. \tag{6}$$

Equation 6 suggests that a test of the $\bar{x}_i$ in a linear model that controls for $x_{ti}$ is actually a specification test for the random effects model. However, such a test will generally be inaccurate because it assumes that all variation between persons is accounted for by $\bar{x}_i$. This problem is readily solved by the insertion of the random effect, yielding

$$\begin{aligned} y_{ti} &= \alpha + \beta_b \bar{x}_i + \beta(x_{ti} - \bar{x}_i) + u_i + e_{ti} \\ &= \alpha + (\beta_b - \beta)\bar{x}_i + \beta x_{ti} + u_i + e_{ti}. \end{aligned} \tag{7}$$

So testing the coefficient for $\bar{x}_i$ when $x_{ti}$ is controlled within the random effects model tests the random effects specification. Although such a specification test is valid, Hausman (1978) suggested a potentially more powerful test of the random effects specification. Neuhaus and McCulloch (2006) show that the decomposition of the association between $x$ and $y$ as shown in equation 7 not only removes bias in estimating $\beta$ in linear models but also greatly reduces bias in generalized linear models with random effects. Such models are useful for binary outcomes, counts, waiting times, and other limited dependent variables where nonlinear functions relate the mean of the outcome to a linear model.

In sum, we now have reviewed two ways to identify the within-person coefficient $\beta$: use a fixed effects model, which can be implemented by centering the outcome and predictor as in equation 4, or use a random effects model in which we control $\bar{x}_i$ to "protect" $u_i$ from an association with $x_{ti}$ (see Raudenbush and Bryk 2002, chapter 5, for a thorough discussion and examples) as in equation 7.

The left-hand side of equation 7, however, suggests a third way: in the model $y_{ti} = \alpha + \beta_b \bar{x}_i + \beta(x_{ti} - \bar{x}_i) + u_i + e_{ti}$, it is really not essential to control $\bar{x}_i$ because $\bar{x}_i$ is orthogonal to $(x_{ti} - \bar{x}_i)$. The third way to identify $\beta$, then, is to estimate a random effects model in which the predictor $x_{ti}$ is adaptively centered (that is, centered cluster by cluster), in this case around $\bar{x}_i$:

$$y_{ti} = \alpha + \beta(x_{ti} - \bar{x}_i) + u_i + e_{ti}. \tag{8}$$

One may reason that controlling for $\bar{x}_i$ as in equation 7 does not hurt and may help, so why eliminate $\bar{x}_i$ as in equation 8? Controlling for $\bar{x}_i$ is reasonable as long as the model includes one or just a few covariates, $x$. However, in many applications the number of covariates will increase, and it often becomes burdensome to control multiple means because these means may be highly intercorrelated. The inclusion of $\bar{x}_i$ in the model is, however, always an option and can easily be applied using the methods described below.

The simple example in the next section suggests that inferences about $\beta$ in equation 8 based on normal distribution and maximum likelihood are in fact identical to those based on the fixed effects model 1 in the case of balanced data. For unbalanced data, the point estimates are identical, while the standard errors may differ negligibly in small samples. Section 4 provides a technical justification for this approach and shows how the concept of adaptive centering can be generalized to research designs with the complex forms of nesting that typically arise in large-scale studies in education.

## 3. A SIMPLE ILLUSTRATIVE EXAMPLE

To illustrate the adaptive centering approach, consider the hypothetical data set in table 1. We have twenty children, $I = 1, \ldots, n = 20$ (rows of the table),

**Table 1.** Outcome Data for 20 Hypothetical Children by 9 Teachers Nested with 3 Schools

| | | School 1 | | | School 2 | | | School 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Teacher | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| | x | −1 | 0 | 1 | −1 | 0 | 1 | −1 | 0 | 1 |
| w | Child | | | | | | | | | |
| 0 | 1 | | | −2.4102 | | | 2.4628 | | | 6.2245 |
| 1 | 2 | | | 3.6396 | | 4.1441 | | | | 11.0898 |
| 1 | 3 | | 2.1827 | | | | 10.1339 | | | 12.3134 |
| 0 | 4 | | | −3170 | | | 3.6596 | | 4.8397 | |
| 0 | 5 | | −.0727 | | | 1.6280 | | | 6.0525 | |
| 0 | 6 | | −2.7852 | | | 1.4795 | | | | 10.0131 |
| 0 | 7 | | .2350 | | | | 6.0839 | | 7.5142 | |
| 0 | 8 | −.8803 | | | 3.5167 | | | | | 9.7337 |
| 0 | 9 | −1.5147 | | | | 5.8636 | | | | 10.2860 |
| 0 | 10 | | | 2.6814 | | | 7.6954 | | | 10.0192 |
| 1 | 11 | 4.4966 | | | 9.5578 | | | 11.1152 | | |
| 1 | 12 | 4.7195 | | | 8.2204 | | | | 14.6855 | |
| 1 | 13 | 4.3609 | | | | | 12.6474 | | 16.8547 | |
| 1 | 14 | 4.7778 | | | | 11.9663 | | | | 18.3998 |
| 1 | 15 | | 8.5264 | | | 12.9066 | | | | 18.6272 |
| 1 | 16 | | 8.6820 | | 11.8265 | | | | 17.0661 | |
| 1 | 17 | | 9.5595 | | | 13.8078 | | | 16.3071 | |
| 1 | 18 | 5.6075 | | | 12.7943 | | | | | 21.075 |
| 1 | 19 | 8.9094 | | | | 13.5301 | | | | 20.049 |
| 0 | 20 | 6.3465 | | | 7.3268 | | | 11.5147 | | |

each observed on three occasions ($T = 3$), with one occasion in each of three schools, $k = 1, 2$, $K = 3$ (see the three major columns). Nested within each school are three teachers, $j = 1, \ldots, J_k = 3$, so there are nine teachers over all. The treatment variable $x \in \{-1, 0, 1\}$ is a teacher characteristic, though it varies within children as they move across schools. This simple data set has the basic structure of the data used in many important educational policy studies reviewed above: students flow across teachers and schools over time; as they do, they encounter varied "treatments" (our $x$). The goal is to exploit the longitudinal character of the data such that within-student variations in treatment are linked to within-student variations in the outcome.

I generated these data in such a way that the assumptions of the conventional random effects model would be violated. Specifically, the outcome

variable depends strongly on characteristics of children and of schools that the analyst cannot observe and that are correlated with the treatment variable $x$. More specifically, I generated the data according to the model

$$y_{tijk} = \theta + \beta x_{tik} + u_i + s_k + \varepsilon_{tijk} \tag{9}$$

where

$$u_i = \gamma w_i + \phi(childid)_i$$

$$s_k = \delta(schoolid - 2)_k$$

with

$$\theta = 0; \quad \beta = 2; \quad \gamma = 5; \quad \delta = 4; \quad \phi = .5; \quad \varepsilon_{tijk} \sim N(0, 1).$$

The errors $\varepsilon$ are mutually independent and independent of the other elements of the model.

In this scenario, $w_i$ is unobserved, and the researcher is unaware of the fact that linear functions of *childid* and *schoolid* contribute to the outcome. The central aim is to estimate $\beta = 2$ using the observed $y$, $x$, *childid*, *schoolid*. The fact that child and school effects are correlated with treatment $x$ invalidates the assumption of the standard random effects model when $u_i$, $s_k$ are regarded as random—that is,

$$\begin{aligned} E(y_{tijk} \mid x_{tik}) &= \theta + \beta x_{ti} + E(u_i + s_k \mid x_{tik}) \\ &= \theta + \beta x_{tik} + E(u_i + s_k) \\ &= \theta + \beta x_{tik}. \end{aligned} \tag{10}$$

The failure of assumption 10 implies that estimation of the random effects model will produce a biased estimate of $\beta$.

### One-Dimensional Confounding

#### *One-Dimensional Fixed Effects Model*

Suppose first that the analyst wishes to control for time-invariant child differences but ignores the possibility of time-invariant school-level confounding. Therefore this analyst fits model 1, yielding the estimates in table 2. The estimate $\hat{\beta} = 5.498, se = 0.856$ is far off the mark of $\beta = 2$, reflecting the failure to control for school-level confounding.

#### *Adaptive Centering with Random Effects*

As an alternative, consider the random effects model (equation 8) that uses adaptive centering of $x$ around the student mean. We add normality

**Table 2.** One-Dimensional Control: OLS Fixed Child Effects

$y_{tijk} = \theta + \beta x_{tik} + u_i + \varepsilon_{tijk}, \varepsilon_{tijk} \sim N(0, \sigma^2)$,

$u_i, i = 1, \ldots, 19$ *fixed*

**Estimates of Fixed Effects**

| Parameter | Estimate | Std. Error | t | Sig. |
|---|---|---|---|---|
| Intercept | 13.894087 | 2.217045 | 6.267 | .000 |
| **x** | **5.498095** | **.865904** | **6.350** | **.000** |
| [childid=1.00] | −17.299841 | 3.366029 | −5.140 | .000 |
| [childid=2.00] | −11.268353 | 3.227033 | −3.492 | .001 |
| [childid=3.00] | −9.349477 | 3.227033 | −2.897 | .006 |
| [childid=4.00] | −14.832045 | 3.227033 | −4.596 | .000 |
| [childid=5.00] | −11.358169 | 3.013434 | −3.769 | .001 |
| [childid=6.00] | −12.825538 | 3.108690 | −4.126 | .000 |
| [childid=7.00] | −11.115732 | 3.108690 | −3.576 | .001 |
| [childid=8.00] | −9.770723 | 3.013434 | −3.242 | .002 |
| [childid=9.00] | −9.015820 | 3.013434 | −2.992 | .005 |
| [childid=10.00] | −12.593491 | 3.366029 | −3.741 | .001 |
| [childid=11.00] | −.006149 | 2.886346 | −.002 | .998 |
| [childid=12.00] | −1.020260 | 2.900742 | −.352 | .727 |
| [childid=13.00] | −.773729 | 2.943507 | −.263 | .794 |
| [childid=14.00] | −2.179455 | 3.013434 | −.723 | .474 |
| [childid=15.00] | −2.373398 | 3.108690 | −.763 | .450 |
| [childid=16.00] | .463474 | 2.943507 | .157 | .876 |
| [childid=17.00] | −.669300 | 3.013434 | −.222 | .825 |
| [childid=18.00] | 1.097582 | 2.943507 | .373 | .711 |
| [childid=19.00] | .268870 | 3.013434 | .089 | .929 |
| [childid=20.00] | 0(a) | 0 | . | . |

**Estimates of Covariance Parameters**

| Parameter | Estimate |
|---|---|
| $\sigma^2$ | 12.496491 |

assumptions $u_i \sim N(0, \tau^2)$, $\varepsilon_{ti} \sim N(0, \sigma^2)$ to facilitate maximum likelihood estimation (table 3). Inferences regarding $\beta$ are identical to those based on the fixed effects model (table 2) with no centering of $x$. A question of interest in the next section will involve when and why these equivalences will hold.

**Table 3.** One-Dimensional Control: Child Random Effects with Person-Mean Centered x

$$y_{tijk} = \theta + \beta(x_{t_{ik}} - \bar{x}_i) + u_i + \varepsilon_{tijk}, \varepsilon_{tijk} \sim N(0, \sigma^2),$$

$$u_i \sim N(0, \tau^2)$$

Note that this gives the same coefficient, standard error, and residual variance estimate as the student fixed effects model.

**Estimates of Fixed Effects**

| Parameter | Estimate | Std. Error | df | t | Sig. |
|---|---|---|---|---|---|
| Intercept | 8.029549 | .927088 | 19 | 8.661 | .000 |
| $(x_{t_{ij_k}} - \bar{x}_i)$ | **5.498095** | **.865904** | **39.000** | **6.350** | **.000** |

**Estimates of Covariance Parameters**

| Parameter | Estimate |
|---|---|
| $\sigma^2$ | 12.496491 |
| $\tau^2$ | 13.024353 |

## Two-Dimensional Confounding

### *Two-Dimensional Fixed Effects Model*

Now suppose that the analyst decides to control for time-invariant school-level differences as well as for time-invariant child differences using the two-dimensional fixed effects model. The estimates are given in table 4. We see that $\hat{\beta} = 2.57, se = 0.288$ is now within the vicinity of the true $\beta = 2$, reflecting the benefit of controlling for unobserved school-level confounding in addition to unobserved student-level confounding.

### *Adaptive Centering with Random Effects*

Now consider the alternative random effects model with two dimensional centering:

$$y_{tijk} = \theta + \beta(x_{tik} - \bar{x}_i - \bar{x}_k + \bar{x}) + u_i + s_k + \varepsilon_{tijk} \tag{11}$$

where

$$u_i \sim N(0, \tau^2), s_k \sim N(0, \omega^2), \varepsilon_{tijk} \sim N(0, \sigma^2),$$

$$\bar{x}_i = \sum_{t=1}^{3} x_{tik}/3, \bar{x}_k = \sum_{i=1}^{20} x_{tik}/20.$$

Equation 11 is a "cross-classified random effects model" (Raudenbush and Bryk 2002, chapter 12) in which time series observations are regarded as crossed by random levels of students and schools. Inferences (table 5) based on maximum

**Table 4.** Two-Dimensional Controls: OLS Fixed Child and School Effects

$y_{tijk} = \theta + \beta x_j + u_i + s_k + \varepsilon_{tijk}, \varepsilon_{tijk} \sim N(0, \sigma^2)$,

$u_i, i = 1, \ldots, 19$ *fixed*

$s_k = 1, 2$ *fixed*

**Estimates of Fixed Effects**

| Parameter | Estimate | Std. Error | df | T | Sig. |
|---|---|---|---|---|---|
| Intercept | 14.642231 | .630345 | 37 | 23.229 | .000 |
| **X** | **2.573106** | **.287937** | **37** | **8.936** | **.000** |
| [childid=1.00] | −11.449864 | .998365 | 37 | −11.469 | .000 |
| [childid=2.00] | −6.393372 | .946257 | 37 | −6.756 | .000 |
| [childid=3.00] | −4.474496 | .946257 | 37 | −4.729 | .000 |
| [childid=4.00] | −9.957064 | .946257 | 37 | −10.523 | .000 |
| [childid=5.00] | −8.433180 | .864876 | 37 | −9.751 | .000 |
| [childid=6.00] | −8.925554 | .901385 | 37 | −9.902 | .000 |
| [childid=7.00] | −7.215747 | .901385 | 37 | −8.005 | .000 |
| [childid=8.00] | −6.845734 | .864876 | 37 | −7.915 | .000 |
| [childid=9.00] | −6.090831 | .864876 | 37 | −7.042 | .000 |
| [childid=10.00] | −6.743514 | .998365 | 37 | −6.755 | .000 |
| [childid=11.00] | −.006149 | .815539 | 37 | −.008 | .994 |
| [childid=12.00] | −.045263 | .821167 | 37 | −.055 | .956 |
| [childid=13.00] | 1.176263 | .837825 | 37 | 1.404 | .169 |
| [childid=14.00] | .745534 | .864876 | 37 | .862 | .394 |
| [childid=15.00] | 1.526586 | .901385 | 37 | 1.694 | .099 |
| [childid=16.00] | 2.413467 | .837825 | 37 | 2.881 | .007 |
| [childid=17.00] | 2.255688 | .864876 | 37 | 2.608 | .013 |
| [childid=18.00] | 3.047574 | .837825 | 37 | 3.637 | .001 |
| [childid=19.00] | 3.193858 | .864876 | 37 | 3.693 | .001 |
| [childid=20.00] | 0(a) | 0 | . | . | . |
| [schoolid=1.00] | −7.679293 | .367143 | 37 | −20.916 | .000 |
| [schoolid=2.00] | −3.340106 | .347120 | 37 | −9.622 | .000 |
| [schoolid=3.00] | 0(a) | 0 | . | . | . |

**Estimates of Covariance Parameters**

| Parameter | Estimate |
|---|---|
| $\sigma^2$ | .997655 |

**Table 5.** Two-Dimensional Controls: Random Child and School Effects with Interaction-Contrast Centering

$y_{tijk} = \theta + \beta x_{tik} + u_i + s_k + \varepsilon_{tik}$,

$\varepsilon_{tik} \sim N(0, \sigma^2)$

$u_i \sim N(0, \tau^2)$,

$s_k \sim N(0, \psi^2)$

**Estimates of Fixed Effects**

| Parameter | Estimate | Std. Error | t | Sig. |
|---|---|---|---|---|
| Intercept | 8.029463 | 2.851520 | 2.816 | .083 |
| $x_{t_{ik}} - \bar{x}_i - \bar{x}_k + \bar{x}$ | **2.573106** | **.287937** | **8.936** | **.000** |

**Estimates of Covariance Parameters**

| Parameter | Estimate |
|---|---|
| $\sigma^2$ | .997655 |
| $\tau^2$ | 16.857298 |
| $\psi^2$ | 21.815022 |

likelihood regarding $\beta$ are identical to those based on the two-dimensional fixed effects model (table 4) with no centering of $x$.

## A Richer Class of Models

The results of this hypothetical example suggest that adaptive centering of treatment indicators with random effects can replicate the fixed effects estimates in any dimension. In fact, within the random effects framework, a richer class of models can be estimated.

### *Accounting for Uncertainty*

For example, it is possible to estimate a random effect for each teacher in the context of our example:

$$y_{tijk} = \theta + \beta(x_{tik} - \bar{x}_i - \bar{x}_k + \bar{x}) + u_i + c_{j(k)} + s_k + \varepsilon_{tijk} \qquad (12)$$

where we add the additional classroom random effect $c_{j(k)} \sim N(0, \psi^2)$. Specification of this random effect allows the analysis to incorporate uncertainty associated with classrooms, presumably providing more realistic standard errors than when such clustering is ignored. An alternative method for obtaining consistent standard errors is the Huber-White approach (White 1980). However, that approach would require multiplying the vector of residuals of every student with the vectors of residuals of every other student, given that the mobility

of students over schools and teachers induces covariances of residuals across all students. This would be computationally difficult in large applications.

#### *Heterogeneous Treatment Effects*

It is straightforward within the random effects framework to allow random coefficients. Consider the model

$$
\begin{aligned}
y_{tijk} &= \theta + u_{0i} + s_{0k} + c_{0j(k)} \\
&\quad + (\beta + u_{1_{ik}} + s_{1k})(x_{t_{ik}} - \bar{x}_i - \bar{x}_k + \bar{x}) + \varepsilon_{tijk} \\
\begin{bmatrix} u_{0i} \\ u_{1_{ik}} \end{bmatrix} &\sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right] \\
\begin{bmatrix} s_{0k} \\ s_{1_k} \end{bmatrix} &\sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \omega_{00} & \omega_{01} \\ \omega_{10} & \omega_{11} \end{pmatrix}\right] \\
c_{j(k)} &\sim N(0, \psi^2) \\
\varepsilon_{tijk} &\sim N(0, \sigma^2).
\end{aligned}
\tag{13}
$$

The variance components $\tau_{11}, \omega_{11}$ parameterize the heterogeneity of the treatment effect across children and schools.

#### *Multilevel Factorial Designs*

We can also readily extend the random effects approach to incorporate multilevel treatments and cross-level interactions. Consider the case of a between-school characteristic or treatment that interacts with $x$:

$$
\begin{aligned}
y_{tijk} &= \theta + u_{0i} + s_{0k} + c_{0j(k)} + \left(\beta + u_{1_{ik}} + s_{1k}\right)\left(x_{t_{ik}} - \bar{x}_i - \bar{x}_k + \bar{x}\right) \\
&\quad + \gamma_0 w_k + \gamma_1 w_k * \left(x_{t_{ik}} - \bar{x}_i - \bar{x}_k + \bar{x}\right) + \varepsilon_{tijk} \\
\begin{bmatrix} u_{0i} \\ u_{1_{ik}} \end{bmatrix} &\sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right] \\
\begin{bmatrix} s_{0k} \\ s_{1_k} \end{bmatrix} &\sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \omega_{00} & \omega_{01} \\ \omega_{10} & \omega_{11} \end{pmatrix}\right] \\
c_{j(k)} &\sim N(0, \psi^2) \\
\varepsilon_{tijk} &\sim N(0, \sigma^2).
\end{aligned}
\tag{14}
$$

In this case $\gamma_0$ is the main effect of the between-school predictor $w$ and $\gamma_1$ is the cross-level interaction effect. Such specifications are not possible within the fixed effects framework.

## 4. GENERAL STATISTICAL APPROACH

Let us now consider a general linear mixed model that can incorporate any number of levels. For example, we might have students within schools within districts in a cross-sectional study or repeated measures on students who are cross classified by teachers within schools within districts. My aim is to derive a general adaptive centering estimator that can remove confounding from one or more sources of nesting, as is done in fixed effects models, while also accurately reflecting the multilevel structure of the data and affording the flexible features of the random effects model. After doing so, I will show the form these estimators take in the case of one dimension of confounding and two dimensions of confounding.

### The General Model

Let us then consider the linear model

$$\mathbf{Y} = \mathbf{1}\theta + \tilde{\mathbf{X}}\boldsymbol{\beta} + \mathbf{A}\mathbf{b} + \mathbf{e}, \quad \mathbf{b} \sim N(\mathbf{0}, \boldsymbol{\Omega}), \quad \mathbf{e} \sim N(\mathbf{0}, \mathbf{V}^*), \tag{15}$$

where $\mathbf{Y}$ is a vector of outcomes; $\theta$ is a fixed, unknown intercept; $\boldsymbol{\beta}$ is a vector of unknown fixed regression coefficients; $\mathbf{b}$ and $\mathbf{e}$ are vectors of unknown random effects; $\mathbf{1}$ is a column vector with every element equal to unity; $\tilde{\mathbf{X}}$ and $\mathbf{A}$ are known design matrices dimensioned conformably, where $\mathbf{A}$ is composed of elements 1 or 0 such that each element of the random effect vector $\mathbf{b}$ is assigned the correct unit; and $\boldsymbol{\Omega}$ and $\mathbf{V}^*$ are positive definite covariance matrices, considered known. The assumption that the covariance components are known is not realistic, but we will be interested here in asymptotic properties; as the covariance component estimators converge to the true parameters, the results in this section will hold approximately. The covariance matrix of the outcome is given by

$$Var(\mathbf{Y}) = \mathbf{V} = \mathbf{A}\boldsymbol{\Omega}\mathbf{A}^T + \mathbf{V}^*. \tag{16}$$

Without loss of generality, we shall center the covariates in $\tilde{\mathbf{X}}$ around their sample means:

$$\mathbf{X} = \tilde{\mathbf{X}} - \mathbf{1}(\mathbf{1}^T\mathbf{V}^{-1}\mathbf{1})^{-1}\mathbf{1}^T\mathbf{V}^{-1}\tilde{\mathbf{X}}. \tag{17}$$

As a result $\mathbf{1}^T\mathbf{V}^{-1}\mathbf{X} = \mathbf{0}$, and the maximum likelihood estimator of the coefficients of interest has the familiar generalized least squares form

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}. \tag{18}$$

(In practice, "grand mean centering" of $\tilde{\mathbf{X}}$ will not be needed because adaptive centering as described below will insure that the overall sample mean of the

covariate matrix will be null. However, grand mean centering as equation 17 simplifies the exposition below.)

A key assumption for this random effects model is that of no association between the random effects **b** or **e** and the predictors, **X**, in which case

$$\begin{aligned} E(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) &= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}E[\mathbf{Y} \mid \mathbf{X}] \\ &= \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}[\mathbf{A}E(\mathbf{b} \mid \mathbf{X}) + E(\mathbf{e} \mid \mathbf{X})] \\ &= \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}[\mathbf{A}E(\mathbf{b}) + E(\mathbf{e})] \\ &= \boldsymbol{\beta}. \end{aligned} \tag{19}$$

We may be willing to stipulate independence of **e** and **X** so that $E(\mathbf{e} \mid \mathbf{X}) = E(\mathbf{e}) = \mathbf{o}$ but not the independence of **b** and **X**. So equation 19 is not generally applicable.

***Definition***

Let us define the adaptively centered design matrix

$$\mathbf{X}^* = \mathbf{X} - \mathbf{A}(\mathbf{A}^T\mathbf{V}^{*-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{V}^{*-1}\mathbf{X} \tag{20}$$

and reformulate our model 15 as

$$\mathbf{Y} = \mathbf{1}\theta + \mathbf{X}^*\boldsymbol{\beta} + \mathbf{A}\mathbf{b} + \mathbf{e}, \quad \mathbf{b} \sim N(\mathbf{o}, \boldsymbol{\Omega}), \quad \mathbf{e} \sim N(\mathbf{o}, \mathbf{V}^*). \tag{21}$$

***Theorem***

The maximum likelihood estimator of $\beta$ and its covariance matrix will be

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{*T}\mathbf{V}^{*-1}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{V}^{*-1}\mathbf{Y}, \quad Var(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^{*T}\mathbf{V}^{*-1}\mathbf{X}^*)^{-1}, \tag{22}$$

where $E(\hat{\boldsymbol{\beta}} \mid \mathbf{X}^*) = \boldsymbol{\beta}$ even if **X** is correlated with **b**.

***Proof***

We know that $\mathbf{V}^{-1} = \mathbf{V}^{*-1} - \mathbf{V}^{*-1}\mathbf{A}(\mathbf{A}^T\mathbf{V}^{*-1}\mathbf{A} + \boldsymbol{\Omega}^{-1})^{-1}\mathbf{A}^T\mathbf{V}^{*-1}$ (Lindley and Smith 1972). Therefore

$$\begin{aligned} \hat{\boldsymbol{\beta}} = {} & [\mathbf{X}^{*T}\mathbf{V}^{*-1}\mathbf{X} - \mathbf{X}^{*T}\mathbf{V}^{*-1}\mathbf{A}(\mathbf{A}^T\mathbf{V}^{*-1}\mathbf{A} + \boldsymbol{\Omega}^{-1})^{-1}\mathbf{A}^T\mathbf{V}^{*-1}\mathbf{X}]^{-1} \\ & * \mathbf{X}^{*T}\mathbf{V}^{*-1}\mathbf{Y} - \mathbf{X}^{*T}\mathbf{V}^{*-1}\mathbf{A}(\mathbf{A}^T\mathbf{V}^{*-1}\mathbf{A} + \boldsymbol{\Omega}^{-1})^{-1}\mathbf{A}^T\mathbf{V}^{*-1}\mathbf{Y}. \end{aligned} \tag{23}$$

However, given the adaptive centering definition (20),

$$\mathbf{A}^T\mathbf{V}^{*-1}\mathbf{X}^* = \mathbf{A}^T\mathbf{V}^{*-1}[\mathbf{X} - \mathbf{A}(\mathbf{A}^T\mathbf{V}^{*-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{V}^{*-1}\mathbf{X}] = \mathbf{o}$$

Hence equation 22 will hold and

$$E(\hat{\boldsymbol{\beta}} \mid \mathbf{X}^*) = \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{V}^{*-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{*-1}\mathbf{A}\,E(\mathbf{b} \mid \mathbf{X}^*) = \boldsymbol{\beta}. \tag{24}$$

I will show in the next sections how to solve these equations in the case of one-dimensional confounding (e.g., confounding of time-invariant child effects) and two-dimensional confounding (e.g., time-invariant child and school effects). The approach extends, of course, to multiple dimensions of confounding. We consider an arbitrary number of levels of nesting and crossing.

### One-Dimensional Confounding

Let us now consider the case in which there is an $L$-level nested structure with a treatment variable defined at level $L$-1. The aim is to remove confounding associated with level $L$ of the model. In the conventional random effects model (Mundlak 1978) we have $L = 2$, where the first level ($L = 1$) involves repeated measures nested within the second level ($L = 2$)—for example, children. However, we might have repeated cross sections of students (level 1) who are nested within classrooms (level 2) within waves of data collection (level 3) within schools (level 4) where the treatment varies over time within schools. In this case, $L = 4$ with treatment defined at level 3. After showing the general result for the $L$-level model, I shall illustrate how the procedure works for two-level ($L = 2$) and three-level models ($L = 3$).

#### *L-Level Model*

Without adaptive centering, the model is

$$\mathbf{Y}_r = \mathbf{1}_r\theta + \mathbf{X}\boldsymbol{\beta} + \mathbf{1}_r\mathbf{b}_r + \mathbf{e}_r, \quad \mathbf{b}_r \sim N(\mathbf{0}, \omega^2\mathrm{I}_r), \quad \mathbf{e}_r \sim N(\mathbf{0}, \mathbf{V}_{L-1,r}). \tag{25}$$

Here $\mathbf{Y}_r$ is the $n_r$ by 1 vector of outcomes within $L$-level unit $r$ having elements $\{y_{ijk\ldots r}\}$. The regression coefficient vector $\boldsymbol{\beta}$ is of central interest. Correspondingly, $\mathbf{X}_r$ is the known design matrix for level-$L$ unit $r$. In the case of one-dimensional blocking, the random effects design is simply the $n_r$ by 1 vector $\mathbf{1}_r$, all of whose elements are unity. The covariance matrix $\mathbf{V}_{L\text{-}1,r} = Var(\mathbf{Y}_r \mid \mathbf{b}_r)$ and $\mathbf{V}_{L,r} = Var(\mathbf{Y}_r) = \omega^2\mathbf{1}_r\mathbf{1}_r^T + \mathbf{V}_{L-1,r}$. We now apply adaptive centering (equation 20), yielding

$$\mathbf{X}_r^* = \mathbf{X}_r - 1_r\bar{\mathbf{X}}_r, \tag{26}$$

where $\bar{\mathbf{X}}_r = (\mathbf{1}_r^T\mathbf{V}_{L-1,r}^{-1}\mathbf{1}_r)^{-1}\mathbf{1}_r^T\mathbf{V}_{L-1,r}^{-1}\mathbf{X}_r$. This will produce an unbiased estimator of $\boldsymbol{\beta}$ because $\mathbf{1}_r^T\mathbf{V}_{L,r}^{-1}\mathbf{X}_r^* = \mathbf{0}$.

#### *Two-Level Model*

The two-level model is a special case of equation 25 with $L = 2$, $r = j$, so that $\mathbf{Y}_j$ is the $n_j$ by 1 vector of outcomes within level-2 unit $j$ having elements

$\{y_{ij}\}$. The covariance matrix $\mathbf{V}_{1,j} = Var(\mathbf{Y}_j \mid \mathbf{b}_j) = \sigma^2\mathbf{I}_j$ and $\mathbf{V}_{2,j} = Var(\mathbf{Y}_j) = \omega^2\mathbf{1}_j\mathbf{1}_j^T + \sigma^2\mathbf{I}_j$ with $\sigma^2$ denoting the level-1 variance. For example, $j$ may denote the person and $i$ the time point; $\sigma^2$ is the within-person variance; and $\omega^2$ is the between-person variance. Without centering, the maximum likelihood estimator is

$$\hat{\boldsymbol{\beta}} = \left(\sum_{j=1}^{J}\mathbf{X}_j^T\mathbf{V}_{2,j}^{-1}\mathbf{X}_r\right)^{-1}\sum_{j=1}^{J}\mathbf{X}_j^T\mathbf{V}_{2,j}^{-1}\mathbf{Y}_r. \tag{27}$$

We now apply adaptive centering equation 20, yielding in this case $\mathbf{X}_j^* = \mathbf{X}_j - 1_j\bar{\mathbf{X}}_j$, where

$$\bar{\mathbf{X}}_j = \left(\mathbf{1}_j^T\mathbf{1}_j\right)^{-1}\mathbf{1}_j^T\mathbf{X}_j = 1/n_j\sum_{i=1}^{n_j}\mathbf{X}_{ij} \tag{28}$$

is simply the arithmetic mean. We then have $\mathbf{1}_j^T\mathbf{X}_j^* = \mathbf{0}$ so that

$$\begin{aligned}\sum_{j=1}^{J}\mathbf{X}_j^{*T}\mathbf{V}_{2,j}^{-1}\mathbf{X}_r^* &= \sigma^{-2}\sum_{j=1}^{J}\mathbf{X}_j^{*T}\left(\omega^2\mathbf{1}_j\mathbf{1}_j^T + \sigma^2\mathbf{I}_j\right)^{-1}\mathbf{X}_j^* \\ &= \sigma^{-2}\sum_{j=1}^{J}\mathbf{X}_j^{*T}\left[\mathbf{I}_j - \mathbf{1}_j\left(\mathbf{1}_j^T\mathbf{1}_j + \sigma^2\omega^{-2}\right)\mathbf{1}_j^T\right]\mathbf{X}_j^* \\ &= \sigma^{-2}\left[\sum_{j=1}^{J}\mathbf{X}_j^{*T}\mathbf{X}_j^* - \sum_{j=1}^{J}\mathbf{X}_j^{*T}\mathbf{1}_j\left(\mathbf{1}_j^T\mathbf{1}_j + \sigma^2\omega^{-2}\right)\mathbf{1}_j^T\mathbf{X}_j^*\right] \\ &= \sigma^{-2}\sum_{j=1}^{J}\mathbf{X}_j^{*T}\mathbf{X}_j^*. \end{aligned} \tag{29}$$

Similarly, $\sum_{j=1}^{J}\mathbf{X}_j^{*T}\mathbf{V}_{2,j}^{-1}\mathbf{Y}_r = \sigma^2\sum_{j=1}^{J}\mathbf{X}_j^{*T}\mathbf{Y}_j$, so in this case we have the OLS estimator

$$\hat{\boldsymbol{\beta}} = \left(\sum_{j=1}^{J}\mathbf{X}_j^T\mathbf{X}_j\right)^{-1}\sum_{j=1}^{J}\mathbf{X}_j^T\mathbf{Y}_j, \quad Var(\hat{\boldsymbol{\beta}}) = \sigma^2\left(\sum_{j=1}^{J}\mathbf{X}_j^{*T}\mathbf{X}_j^*\right)^{-1}. \tag{30}$$

#### *Three-Level Model*

The three-level model is a special case of equation 25, with $L = 3$ and $r = k$, where $\mathbf{Y}_k$ is the $n_k$ by 1 vector of outcomes within level-3 unit $k$ having elements $\{y_{ijk}\}$. The covariance matrix is $Var(\mathbf{Y}_k) = \mathbf{V}_{3,k} = \omega^2\mathbf{1}_k\mathbf{1}_k^T + \mathbf{V}_{2k}$, with $\mathbf{V}_{2,k} = \bigoplus_{j=1}^{J_k}\mathbf{V}_{2jk} = Var(\mathbf{Y}_k \mid \mathbf{b}_k)$, and $\mathbf{V}_{2jk} = \tau^2\mathbf{1}_{jk}\mathbf{1}_{jk}^T + \sigma^2\mathbf{I}_{jk}$, where $\sigma^2$ is the

level-1 variance, $\tau^2$ is the level-2 variance, $\mathbf{1}_{jk}$ is an $n_{jk}$ by 1 vector with all elements equal to unity, and $\mathbf{I}_{jk}$ is the $n_{jk}$ by $n_{jk}$ identity matrix. An example involves students $i = 1, \ldots, n_{jk}$ nested within classrooms $j = 1, \ldots, J_k$ that are in turn nested within schools $k = 1, \ldots, K$.

In the case of $\omega^2$, $\tau^2$, and $\sigma^2$ known, the maximum likelihood estimator of the regression coefficients is

$$\hat{\boldsymbol{\beta}} = \left( \sum_{k=1}^{K} \mathbf{X}_k^T \mathbf{V}_{3,k}^{-1} \mathbf{X}_k \right)^{-1} \sum_{k=1}^{K} \mathbf{X}_k^T \mathbf{V}_{3,k}^{-1} \mathbf{Y}_k. \tag{31}$$

Applying adaptive centering (equation 20), we now set

$$\mathbf{X}_k^* = \mathbf{X}_k - \mathbf{1}_k \bar{\mathbf{X}}_k$$

where

$$\begin{aligned} \bar{\mathbf{X}}_k &= \left( \mathbf{1}_k^T \mathbf{V}_{2,k}^{-1} \mathbf{1}_k \right)^{-1} \mathbf{1}_k^T \mathbf{V}_{2,k}^{-1} \mathbf{X}_k \\ &= \left( \sum_{j=1}^{J_k} (\tau^2 + \sigma^2 / n_{jk})^{-1} \right)^{-1} \sum_{j=1}^{J_k} (\tau^2 + \sigma^2 / n_{jk})^{-1} \bar{\mathbf{X}}_{jk}. \end{aligned} \tag{32}$$

Here $\bar{\mathbf{X}}_{jk}$ is the unweighted average of $\mathbf{X}_{ijk}$.We then have $\mathbf{1}_k^T \mathbf{V}_{2,k}^{-1} \mathbf{X}_k^* = \mathbf{0}$. Substituting $\mathbf{X}_j^*$ for $\mathbf{X}_j$ in equation 31, we now find that

$$\begin{aligned} \sum_{k=1}^{J} \mathbf{X}_k^{*T} \mathbf{V}_{3,k}^{-1} \mathbf{X}_k^* &= \sum_{k=1}^{K} \mathbf{X}_k^{*T} \left[ \mathbf{V}_{2,k}^{-1} - \mathbf{V}_{2,k}^{-1} \mathbf{1}_k \left( \mathbf{1}_k^T \mathbf{V}_{2,k}^{-1} \mathbf{1}_k + \omega^{-2} \mathbf{I}_k \right) \mathbf{1}_k^T \mathbf{V}_{2,k}^{-1} \right] \mathbf{X}_k^* \\ &= \sum_{k=1}^{K} \mathbf{X}_k^{*T} \mathbf{V}_{2,k}^{-1} \mathbf{X}_k^* - \sum_{k=1}^{K} \mathbf{X}_k^{*T} \mathbf{V}_{2,k}^{-1} \mathbf{1}_k (\mathbf{1}_k^T \mathbf{V}_{2,k}^{-1} \mathbf{1}_k + \omega^{-2} \mathbf{I}_k) \mathbf{1}_k^T \mathbf{V}_{2,k}^{-1} \mathbf{X}_k^* \\ &= \sum_{k=1}^{K} \mathbf{X}_k^{*T} \mathbf{V}_{2,k}^{-1} \mathbf{X}_k^*. \end{aligned} \tag{33}$$

Using a similar argument,

$$\sum_{k=1}^{J} \mathbf{X}_k^{*T} \mathbf{V}_{3,k}^{-1} \mathbf{Y}_k = \sum_{k=1}^{K} \mathbf{X}_k^{*T} \mathbf{V}_{2,k}^{-1} \mathbf{Y}_k. \tag{34}$$

With these results in mind, we can see that

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left( \sum_{k=1}^{J} \mathbf{X}_k^{*T} \mathbf{V}_{2,k}^{-1} \mathbf{X}_k^* \right)^{-1} \sum_{k=1}^{K} \mathbf{X}_k^{*T} \mathbf{V}_{2,k}^{-1} \mathbf{Y}_k \\ Var(\hat{\boldsymbol{\beta}}) &= \left( \sum_{k=1}^{J} \mathbf{X}_k^{*T} \mathbf{V}_{2,k}^{-1} \mathbf{X}_k^* \right)^{-1}, \end{aligned} \tag{35}$$

which has the form of a generalized least squares estimator based on a two-level hierarchical linear model.

#### Two-Dimensional Confounding

We now consider the case in which observations are nested within the cells of a two-way cross classification. The idea is to remove the confounding associated with two dimensions. For example, we might have repeated observations on students cross-classified by schools. Alternatively, the time series may be cross classified by students and schools where there are also classrooms nested within schools. Our model is

$$\begin{aligned} &\mathbf{Y} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{R}\mathbf{u} + \mathbf{C}\mathbf{v} + \mathbf{e}, \\ &\mathbf{u} \sim N(\mathbf{o}, \omega^2\mathbf{I}), \quad \mathbf{v} \sim N(\mathbf{o}, \psi^2\mathbf{I}), \quad \mathbf{e} \sim N(\mathbf{o}, \mathbf{V}^*). \end{aligned} \tag{36}$$

Here $\mathbf{R}$ and $\mathbf{C}$ are matrices of indicators that assign random effects $\mathbf{u}$ to the appropriate "rows" (e.g., children) and $\mathbf{v}$ to the appropriate "columns" (e.g., schools), respectively. Equation 36 is a special case of the general model (25) with

$$\begin{aligned} &\mathbf{A} = (\mathbf{R}\ \mathbf{C}), \mathbf{b} = (\mathbf{u}^T\ \mathbf{v}^T)^T \quad \text{and} \\ &\Omega = \begin{bmatrix} \omega^2\mathbf{I}, & \mathbf{o} \\ \mathbf{o} & \psi^2\mathbf{I} \end{bmatrix}. \end{aligned} \tag{37}$$

Adaptive centering requires

$$\mathbf{X}^{*T}\mathbf{V}^{*-1}\mathbf{A} = (\mathbf{X}^{*T}\ \mathbf{V}^{*-1}\mathbf{R}\ \mathbf{X}^{*T}\mathbf{V}^{*-1}\mathbf{C}) = (\mathbf{o}\ \mathbf{o}). \tag{38}$$

This suggests that we regress $\mathbf{x}$ on $\mathbf{C}$ and $\mathbf{R}$, using generalized least squares with weight matrix $\mathbf{V}^{*-1}$, then extract residuals $\mathbf{x}^*$. We illustrate this approach in the case of $\mathbf{V}^* = \sigma^2\mathbf{I}$.

Here is an illustrative example: Marshall Jean at the University of Chicago has assembled a data set on more than two hundred thousand students moving across more than five hundred schools in Chicago. The aim of the study is to estimate the impact of certain school-level characteristics, which, along with a vector of time-varying covariates, are collected in the matrix $\mathbf{X}$. The two-dimensional fixed effects estimation would remove time-varying confounding attributable to students and schools. However, this is a computationally difficult task. Can the adaptive centering approach be feasibly implemented in this case?

In principle, we might regress $\mathbf{X}$ on $\mathbf{C}$ and $\mathbf{R}$, using OLS (given the assumption $\mathbf{V}^* = \sigma^2\mathbf{I}$). We would then extract residuals $\mathbf{X}^*$, achieving condition

(38). But this is computationally demanding given the dimension of **R** (over 200,000 rows). We used the following procedure:

Step 1. Regress **X** on **R**, save the residuals. This is equivalent to centering around the child mean. Specifically, define $\mathbf{x}_{tik}$ as the vector of explanatory variables for student $i$ attending school $k$ at time $t$, $t = 1, \ldots, T_{ik}$; $i = 1, \ldots, n_{jk}$; $k = 1, \ldots, K$. Then we have $\mathbf{x}^*_{tik} = \mathbf{x}_{tik} - \bar{x}_{\cdot i}$.

Step 2. For each observation, regress **C** on **R**, save the residuals. This is easier than it sounds. Simply define dummy variable $C_{tik} = 1$ if student $i$ attends school $k$ at time $t$; $C_{tik} = 0$ otherwise. Do this for each school $k = 1, \ldots K$ so that there are $K$ dummy variables per occasion per student. Now compute $C^*_{tik} = C_{tik} - n_{ik}/n_i$, where $n_{ik}$ is the number of observations for student $i$ in school $k$ and $n_i$ is the total number of observations for student $i$. Thus $n_{ik}/n_i$ is the proportion of student $i$'s observations that occurred while in school $k$ . The collection of those is equivalent to the predicted value of **C** given **R** so that $C^*_{tik}$ are the residuals.

Step 3. Compute a regression with $\mathbf{X}^*_{tik}$ (from step 1) as the outcome and with $J$ predictors $C^*_{tik} = C_{tik} - n_{ik}/n_i$, $k = 1, \ldots K$ (from step 2). Save the residuals from this regression, that is, save $\mathbf{X}^{**}_{tik} = \mathbf{X}^*_{tik} - \hat{E}(\mathbf{X}^*_{tik} \mid C^*_{ti1}, \ldots, C^*_{tiK})$. These in fact are the variables to be used in our regressions. This approach satisfies equation 38, removing time-invariant confounding attributable to children and schools. In the special case of balanced data—that is, when each student is observed the same number of times in each school (as in the hypothetical case of section 3)—the result is $\mathbf{X}^{**}_{tik} = \mathbf{X}_{tik} - \bar{\mathbf{X}}_{\cdot i \cdot} - \bar{\mathbf{X}}_{\cdot \cdot k} + \bar{\mathbf{X}} \ldots$ as in section 3.

The work reported here was supported by funds from the Spencer Foundation for the project "Improving Research on Instruction: Models, Designs, and Analytic Methods." An earlier version of this article was presented at the National Conference on Value-Added Modeling, 22–24 April 2008, at the University of Wisconsin–Madison. I wish to thank Paul Allison, Paul Rathouz, Guanglei Hong, Derek Neal, Sean Reardon, and the editor for their thoughtful comments.

## REFERENCES

Allison, Paul. 2005. *Fixed-effects regression methods for longitudinal data using the SAS system*. Cary, NC: SAS Institute.

Bifulco, Robert, and Helen F. Ladd. 2006. The impacts of charter schools on student achievement: Evidence from North Carolina. *Education Finance and Policy* 1 (1): 50–90.

Cambell, Donald T., and Julian C. Stanley. 1963. *Experimental and quasi-experimental designs for research*. Chicago: Rand-McNally.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2007a. Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review* 24 (4): 377–92.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2007b. How and why do teacher credentials matter for student achievement? NBER Working Paper No. 12828.

Firebaugh, Glenn. 1978. A rule for inferring individual level relationships from aggregate data. *American Sociological Review* 43 (4): 557–72.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2004. Disruption versus Tiebout improvement: The costs and benefits of switching schools. *Journal of Public Economics* 88 (9): 1721–46.

Hanushek, Eric A., and Steven G. Rivkin. 2006. Teacher quality. In *The handbook of the economics of education*, volume 2, edited by Eric A. Hanushek and Finis Welch, pp. 1051–78. Amsterdam: Elsevier.

Harris, Douglas N., and Tim R. Sass. 2008. The effect of National Board–certified teachers on student achievement. *Journal of Public Policy Analysis and Management* 28 (1): 55–80.

Hausman, Jerry A. 1978. Specification tests in econometrics. *Econometrica* 46: 1251–71.

Hong, Guanglei, and Stephen W. Raudenbush. 2008. Causal inference for time-varying instructional treatments. *Journal of Educational and Behavioral Statistics* 33 (3): 333–62.

Jacob, Brian, and Lars Lefgren. 2004a. Remedial education and student achievement: A regression discontinuity analysis. *Review of Economics and Statistics* 86 (1): 226–44.

Jacob, Brian, and Lars Lefgren. 2004b. The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources* 39 (1): 50–79.

Lindley, D. V., and A. F. M. Smith. 1972. *Bayes estimates for the linear model. Journal of the Royal Statistical Society, Series B* 34 (1): 1–41.

Mundlak, Yair. 1978. On the pooling of time series and cross-sectional data. *Econometrica* 46 (1): 69–86.

Neal, Derek, and Diane Whitmore Schanzenbach. 2009. Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics*. In press.

Neuhaus, John M., and Charles E. McCulloch. 2006. Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *Journal of the Royal Statistical Society, Series B* 68 (5): 859–72.

Paterson, Lindsay. 1991. Trends in attainment in Scottish secondary schools. In *Schools, classrooms, and pupils: International studies of schooling from a multilevel perspective*, edited by Stephen W. Raudenbush and J. Douglas Willms, pp. 85–114. San Diego, CA: Academic Press.

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical linear models: Applications and data analysis methods*, 2nd ed. Thousand Oaks, CA: Sage Publications.

Robins, James. 2000. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical models in epidemiology, the environment, and clinical trials*, edited by M. Elizabeth Halloran and Donald Berry, pp. 95–134. New York: Springer.

Roderick, Melissa, and Jenny Nagoka. 2005. Retention under Chicago's high-stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis* 27 (4): 309–40.

White, Halbert. 1980. A heteroscedasticity-consistent covariance matrix estimator and a direct test of heteroscedasticity. *Econometrica* 48: 817–38.

Willms, J. Douglas. 1986. Social class segregation and its relationship to pupils' examination results in Scotland. *American Sociological Review* 55: 224–41.