On the common but problematic specification of conflated random slopes in multilevel models

By

Jason D. Rights

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

May 10, 2019

Nashville, Tennessee

Approved:

Sonya K. Sterba, Ph.D.

Sun-Joo Cho, Ph.D.

Kristopher J. Preacher, Ph.D.

Andrew J. Tomarken, Ph.D.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

Chapter I

Introduction

Multilevel modeling (MLM; also known as linear mixed effects modeling and hierarchical linear modeling) is a popular and useful tool for analyzing nested data structures, such as students nested within schools or repeated observations nested within persons (Demidenko, 2004; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). MLM allows researchers to simultaneously examine the influence of observation-level/ level-1 predictors (e.g., student characteristics) and level-2/ cluster-level predictors (e.g., school characteristics) on an outcome of interest.

In the methodological literature, it has been widely recognized that a level-1 variable can have both a *between-cluster* and a *within-cluster* fixed effect (Algina & Swaminathan, 2011; Cronbach, 1976; Curran et al. 2012; Curran & Bauer, 2011; Enders & Tofighi, 2007; Hedeker & Gibbons, 2006; Hofmann & Gavin, 1998; Preacher, 2011; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). To explain, consider that the observed value of a level-1 variable for a given observation *i* nested within cluster *j* is implicitly the sum of two distinct parts: (a) the overall or aggregate level for cluster *j*, and (b) observation *i*'s deviation from the cluster-aggregated level. These two parts can each exert an influence on a particular outcome; the fixed effect of the former can be termed the between-cluster fixed effect, and that of the latter the within-cluster fixed effect. Importantly, these effects need not be the same nor similar.

As a concrete example wherein the between-cluster and within-cluster fixed effects are different, Baldwin, Wampold, and Imel (2007) examined patients nested within clinicians and predicted patient outcomes from therapeutic alliance, defined as the degree to which the patient-clinician dyad engages in “collaborative, purposive work” (Hatcher & Barends, 2006). With a random-intercept MLM, Baldwin et al. (2007) found that clinicians (i.e., clusters) with a higher mean therapeutic alliance across all patients had lower negative outcomes among their patients; in other words, there was evidence of a *between-cluster* or between-clinician effect. However, they found that, for patients who had the same clinician, therapeutic alliance was not predictive of outcomes; in other words, there was no apparent *within-cluster* or within-clinician effect. Thus, there is evidence that the influence of therapeutic alliance is realized at the clinician-level, and the influence of patient’s individual propensity to engage with clinicians may not be as important. Indeed, as Baldwin et al. (2007) note as an implication of their findings, “it would behoove therapists to attend to their own contributions to the alliance and focus less on characteristics of the patient that impede the development of the alliance (p. 851).”

Because of the possibility that the between-cluster and within-cluster fixed effects may differ, it is widely recommended to explicitly disaggregate these (e.g., Algina & Swaminathan, 2011; Cronbach, 1976; Curran et al. 2012; Curran & Bauer, 2011; Enders & Tofighi, 2007; Hedeker & Gibbons, 2006; Hofmann & Gavin, 1998; Preacher, 2011; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). As an example, prior to the Baldwin et al. (2007) study, researchers studying the therapeutic-alliance/ patient-outcome relationship had not disaggregated and instead obtained estimates of this relationship that were non-zero but, implicitly, a blend of the between-clinician and within-clinician effects; such an estimate that fails to disaggregate the level-specific effects can be said to be *conflated* (e.g., Preacher, 2011). This conflated estimate led many to assume that patient-level differences, rather than clinician-level differences, were responsible for the relationship (e.g., Mallinckrodt, 2000). The results from Baldwin et al. (2007), however, suggest this to be an ecological fallacy (Robinson, 1950) wherein inferences are inappropriately made regarding individuals (patients) based on group (clinician) data. Disaggregation helps researchers to avoid such fallacies and make appropriate, level-specific inferences.

To disaggregate level-specific effects, researchers commonly use contextual effect models (defined shortly). In the methodological literature, the disaggregation afforded by this model has been mainly discussed for fixed slopes in the context of fixed slope models. However, the very same points and explanations about the utility and interpretation of contextual effect models have often been directly applied to *random* slope models—without mention of what complexities this extension might pose regarding disaggregation of level-specific effects (e.g., Algina & Swaminathan, 2011; Hoffman & Stawski, 2009; Hox, 2010; Kreft, de Leeuw, & Aiken, 1995; Paccagnella, 2006; Snijders & Bosker, 2012). Consequently, random-slope contextual effect models have become widely used in practice by researchers interested in disaggregating level-specific effects (e.g., Bliese & Britt, 2001; Deemer, Marks, & Miller, 2017; Diez-Roux, Link, & Northridge, 2000; Espelage, Holt, & Henkel, 2003; Fischer, Rodriguez Mosquera, van Vianen, & Manstead, 2004; Hoffman & Stawski, 2009; Kidwell, Mossholder, & Bennett, 1997; Lee, 2009; Lee & Bryk, 1989; Merlo, Chaix, Yang, Lynch, & Råstam, 2005; Poteat, Espelage, & Green 2007; Schempf & Kaufman, 2012; Titus, 2004).

There are several problems underlying this current situation:

(1) It is not well understood that there are two distinct types of slope conflation that can

occur in MLM: conflation of the *fixed* component of the slope and conflation of the *random* component of the slope. For the slope of a level-1 variable, there can be conflation in one of these components (here called *partial conflation*) or both of these components (here called *full conflation*). Although one of these types of conflation—that of the fixed component—has been widely-recognized and the other type of conflation—that of the random component—has been largely (but not completely) ignored, the exhaustive possibilities—full conflation, partial conflation via the fixed component, and partial conflation via the random component—have not previously been fully enumerated.

(2) In particular, for contextual effect models with random slopes, there is little appreciation that these involve a *conflated random* component of the slope, despite having an *unconflated fixed* component of the slope.

(3) There are negative consequences of fitting a contextual effect model with a conflated random slope that researchers need to understand. Such consequences have not been explained or demonstrated and can include erroneous interpretation and inferences.

(4) Researchers are in need of a full delineation of which random slope MLM specifications yield an unconflated random component, and are in need of recommendations regarding which to use for particular purposes.

This dissertation will address each of these problems. First I review the well-known concept of fixed conflation for a fixed slope MLM with an uncentered (or grand-mean-centered) level-1 predictor. I then review how a fixed-slope contextual effect model disaggregates the fixed component, that is, disentangles the between-cluster fixed effect and within-cluster fixed effect. Next I demonstrate the simultaneous conflation of both the fixed and random component in a random slope MLM with an uncentered (or grand-mean-centered) level-1 predictor. I then show that the conventional random-slope contextual effect model disaggregates the fixed portion, but, importantly, fails to disaggregate the random component, yielding a variance component estimate that conflates level-specific effects. I then discuss and demonstrate the negative consequences of such random conflation, including the potential for erroneous interpretation and inferences pertaining to the estimated slope variance, as well as inappropriate standard errors for fixed effects. Next, noting that the concept of fixed and/or random conflation is important to consider for any MLM with level-1 variables, I provide a general taxonomy for any such model wherein I distinguish model specifications that are fully conflated vs. partially conflated vs. unconflated. For instance, within this taxonomy I show that either group-mean-centered models (defined later) or contextual effect models with random components for the cluster mean (defined later) are unconflated models that fully disaggregate level-specific effects. I then provide evidence via a simulation study that these unconflated models (including the proposed unconflated random-slope contextual effect model) provide better type I error, power, and bias for the random slope variance and provide more accurate standard errors for fixed effects than the widely used conventional random-slope contextual effect model. I provide recommendations for practice in light of these results. Finally, I illustrate the concepts presented with an empirical example and discuss avenues for future research.

Chapter II

Conflation of the fixed slope in a fixed slope model

Uncentered MLM with fixed slopes: Conflation of the fixed component. Consider first a simple random-intercept MLM with a single, uncentered (or grand-mean-centered)[1] level-1 predictor:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij}$$

$$e_{ij} \sim N(0, \sigma^2) \qquad (1)$$

Here, I am modeling a continuous outcome y_{ij} for observation *i* nested within cluster *j*. The level-1 residual, e_{ij}, is normally distributed with a variance of σ^2. The cluster-specific regression coefficients are defined as

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$

$$u_{0j} \sim N(0, \tau_{00}) \qquad (2)$$

The fixed component of the cluster-specific intercept, β_{0j}, is given by γ_{00} and cluster-specific deviations from this mean are given as u_{0j}, which are normally distributed with variance τ_{00}. The fixed slope is given as γ_{10}. This yields the reduced form equation:

---

[1] Note that this level-1 predictor could also be grand-mean-centered by subtracting from x_{ij} its grand mean across all observations. Grand-mean-centering has no impact on whether or not level-specific effects are disaggregated or conflated, and hence I will not focus on the distinction between uncentered and grand-mean-centered predictors. *Group*-mean-centering (subtracting from x_{ij} its cluster mean), on the other hand, will be relevant later in the dissertation.

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + e_{ij} \tag{3}$$

For all models throughout this dissertation, I will focus on such reduced form expressions.

To make the conflation in Equation (3) apparent, I next equivalently express this model by substituting $x_{ij} = x_{ij} - x_{\bullet j} + x_{\bullet j}$. In other words, x_{ij} is replaced by $x_{ij} - x_{\bullet j} + x_{\bullet j}$, which is the sum of its between-cluster portion ($x_{\bullet j}$, the aggregate score for cluster *j*)[2] plus its within-cluster portion ($x_{ij} - x_{\bullet j}$, observation *i*'s deviation from the cluster-aggregated score).

$$\begin{aligned} y_{ij} &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\bullet j} + x_{\bullet j}) + u_{0j} + e_{ij} \\ &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\bullet j}) + \gamma_{10} x_{\bullet j} + u_{0j} + e_{ij} \end{aligned} \tag{4}$$

Here the within-cluster fixed effect (slope of $x_{ij} - x_{\bullet j}$) is γ_{10} and the between-cluster fixed effect (slope of $x_{\bullet j}$) is also γ_{10}. In other words this model constrains the within-cluster fixed effect to be *exactly equal* to the between-cluster fixed effect; this amounts to conflation of the fixed component, which I here term *fixed conflation*.

Such a constraint on the fixed component of the slope is problematic when, in the population, the level-specific fixed effects are not equivalent. Consider again the therapeutic alliance example and suppose that, in Equation (3), y_{ij} represents patient outcomes (e.g., depressive symptoms) and x_{ij} represents therapeutic alliance. If it is actually the case that, as found by Baldwin et al. (2007), the influence on outcomes of a patient's propensity for higher therapeutic alliance is non-existent (or at least negligible) but the clinician's average therapeutic alliance has a negative effect (i.e., patients who see clinicians with better alliance-

[2] In this dissertation, I assume that the cluster-mean is measured without error. In the Discussion section, I address the extension to latent variable models wherein one would use a latent cluster mean rather than an observed cluster mean (Lüdtke et al., 2008).

forming skills have less severe negative outcomes), then the obtained slope estimate for the uncentered model will not accurately reflect the true generating process. In fact, rather than getting a "pure" estimate of either one of these effects, the estimated slope will instead be a weighted average of the two (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Cronbach (1976) first recognized this issue in the social sciences, calling such an estimate an "uninterpretable blend" of the two effects.

Contextual effect model with fixed slopes: Unconflation of the fixed component. To avoid such fixed conflation, a commonly used approach is to fit a contextual effect model (e.g., Boateng, 2016; Enders & Tofighi, 2007; Henry & Slater, 2007; Hoffman & Stawski, 2009; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). This adds to the uncentered model in Equation (3) a fixed slope of $x_{\bullet j}$, and is thus given as

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + \gamma_{01} x_{\bullet j} + u_{0j} + e_{ij}$$

$$e_{ij} \sim N(0, \sigma^2)$$

$$u_{0j} \sim N(0, \tau_{00}) \tag{5}$$

with γ_{01} denoting the *contextual effect*, that is, the fixed effect of $x_{\bullet j}$ controlling for x_{ij}. The contextual effect, γ_{01}, can also be interpreted as the difference between the within-cluster and between-cluster fixed effects of x_{ij}.

To make its disaggregation of the fixed component apparent, I can re-express the fixed-slope contextual effect model as

$$\begin{aligned} y_{ij} &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\bullet j} + x_{\bullet j}) + \gamma_{01} x_{\bullet j} + u_{0j} + e_{ij} \\ &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\bullet j}) + (\gamma_{01} + \gamma_{10}) x_{\bullet j} + u_{0j} + e_{ij} \end{aligned} \tag{6}$$

Here, the within-cluster fixed effect is given as γ_{10}, whereas the between-cluster fixed effect is $\gamma_{01} + \gamma_{10}$. Importantly, unlike the uncentered model in Equation (3), these two fixed slopes are no longer assumed to be equal ($\gamma_{01} + \gamma_{10} \neq \gamma_{10}$); hence the contextual effect model unconflates the fixed component of the slope. Thus, for the therapeutic alliance example, this model could accurately reflect a non-existent within-clinician fixed effect (by having a γ_{10} of 0) and a negative between-clinician fixed effect (by having a γ_{01} that is negative).

Chapter III

Conflation of both the fixed and random components of the slope in a random slope model

Uncentered MLM with random slopes: Conflation of the fixed and random components. Consider next a random slope MLM with a single uncentered (or grand-mean-centered) level-1 predictor. This entails adding to Equation (3) a random component of the slope of x_{ij}:

$$
\begin{gathered}
y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{1j}x_{ij} + u_{0j} + e_{ij} \\
e_{ij} \sim N(0, \sigma^2) \\
\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix}\right)
\end{gathered}
\tag{7}
$$

Here, u_{1j} denotes the cluster-specific deviation from the fixed component of the slope of x_{ij}. The random effects are multivariate normally distributed with a covariance matrix $\mathbf{T}$ that includes the slope variance (τ_{11}) and the intercept-slope covariance (τ_{10}). Re-expressing x_{ij} as the sum of its level-specific parts yields:

$$
\begin{aligned}
y_{ij} &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\bullet j} + x_{\bullet j}) + u_{1j}(x_{ij} - x_{\bullet j} + x_{\bullet j}) + u_{0j} + e_{ij} \\
&= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\bullet j}) + u_{1j}(x_{ij} - x_{\bullet j}) + \gamma_{10}x_{\bullet j} + u_{1j}x_{\bullet j} + u_{0j} + e_{ij}
\end{aligned}
\tag{8}
$$

Here, not only is there again conflation of the fixed component—the fixed component of the slope of $x_{\bullet j}$ (i.e., γ_{10}) equals the fixed component of the slope of $x_{ij} - x_{\bullet j}$ (i.e., γ_{10})—but there is now also conflation of the random component of the slope. Specifically, the random component of the slope of $x_{ij} - x_{\bullet j}$ (i.e., u_{1j}) is exactly equal to the random component of the slope of $x_{\bullet j}$ (i.e., u_{1j}). I term this *random conflation.*

Contextual effect model with random slopes: Unconflation of the fixed component but conflation of the random component. Next I consider the conventional contextual effect model with a random slope of x_{ij}, which is used frequently in applied practice (e.g., Bliese & Britt, 2001; Deemer et al., 2017; Diez-Roux et al., 2000; Espelage et al., 2003; Fischer et al., 2004; Hoffman & Stawski, 2009; Kidwell et al., 1997; Lee, 2009; Lee & Bryk, 1989; Merlo et al., 2005; Poteat et al., 2007; Schempf & Kaufman, 2012; Titus, 2004). From the fixed-slope contextual effect model in Equation (5), I simply add a random slope of x_{ij} to form the conventional random-slope contextual effect model:

$$
\begin{gathered}
y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}x_{\bullet j} + u_{1j}x_{ij} + u_{0j} + e_{ij} \\
e_{ij} \sim N(0, \sigma^2) \\
\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix}\right)
\end{gathered}
\tag{9}
$$

Re-expressing x_{ij} as the sum of its level-specific parts then yields:

$$
\begin{aligned}
y_{ij} &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\bullet j} + x_{\bullet j}) + \gamma_{01}x_{\bullet j} + u_{1j}(x_{ij} - x_{\bullet j} + x_{\bullet j}) + u_{0j} + e_{ij} \\
&= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\bullet j}) + (\gamma_{10} + \gamma_{01})x_{\bullet j} + u_{1j}(x_{ij} - x_{\bullet j}) + u_{1j}x_{\bullet j} + u_{0j} + e_{ij}
\end{aligned}
\tag{10}
$$

This re-expression shows that this conventional random-slope contextual effect model once again unconflates the fixed component of the slope of x_{ij} (since $\gamma_{10} \neq \gamma_{10} + \gamma_{01}$). Unfortunately, however, it *conflates the random component*, in that the residual of the slope of the purely level-1 predictor $x_{ij} - x_{\bullet j}$ (i.e., u_{1j}) is constrained equal to the residual of the slope of the purely level-2 predictor $x_{\bullet j}$ (i.e., u_{1j}). Thus, the contextual effect model with a random slope has an unconflated fixed component but a conflated random component.

The fact that the conventional random-slope contextual effect (or uncentered) MLM yields conflated random components has largely gone unappreciated. It was first briefly noted in an exchange in the *Multilevel Modelling Newsletter* (initially by Raudenbush, 1989; responses by Longford, 1989; Plewis, 1989). But in the intervening 30 years it has rarely been mentioned in the methodological literature (Enders & Tofighi, 2007; Hoffman, 2015; Wu & Wooldridge, 2005), and has been virtually ignored in the applied literature—in contrast to the considerable attention and concern that has been paid to conflation of the fixed component in both the methodological and applied literatures (e.g., Algina & Swaminathan, 2011; Cronbach, 1976; Curran et al. 2012; Curran & Bauer, 2011; Enders & Tofighi, 2007; Hedeker & Gibbons, 2006; Hofmann & Gavin, 1998; Preacher, 2011; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012).

The reason random conflation has gone unappreciated may have to do with the fact that prior authors did not elaborate nor demonstrate the specific issues and implications associated with it. To extend prior work and ensure future researchers can recognize and avoid random conflation, I next explain and show how random conflation leads to erroneous interpretation and inferences regarding the presence and/or degree of slope heterogeneity across clusters, how it yields inappropriate standard errors for fixed effects, and how it can be prevented.

As background for explaining why random conflation is problematic, I must first review the distinction between a random slope of a purely level-1 predictor (which is already well understood) and that of a purely level-2 predictor (which is not widely understood, according to Goldstein [2011] and Snijders & Berkhof [2008]).

A random slope of a purely level-1 predictor $x_{ij} - x_{\bullet j}$ implies that the within-cluster effect of $x_{ij} - x_{\bullet j}$ depends on cluster membership, and thus certain clusters have stronger effects than others. Such a situation is reflected in Figure 1 Panel A, wherein each line represents a cluster-specific regression line of y_{ij} on $x_{ij} - x_{\bullet j}$.

In contrast, a random slope of the purely level-2 predictor $x_{\bullet j}$ reflects *intercept heteroscedasticity* across clusters (e.g., Goldstein, 2011; Rights & Sterba, 2016; Snijders & Bosker, 2012). That is, the conventional random-slope contextual effect model from Equation (9) implies a heteroscedastic intercept variance, here denoted τ_{22j} for cluster *j*, that varies as function of $x_{\bullet j}$. More specifically, the intercept variance can be expressed as a quadratic function[3] of $x_{\bullet j}$ (e.g., Goldstein, 2011; Rights & Sterba, 2016; Snijders & Bosker, 2012), as shown in Appendix A:

$$\begin{aligned} \tau_{22j} &= \mathrm{var}(u_{0j} + u_{1j}x_{\bullet j} \mid x_{\bullet j}) \\ &= \tau_{00} + 2\tau_{10}x_{\bullet j} + \tau_{11}x_{\bullet j}^2 \end{aligned} \tag{11}$$

Such heteroscedastic intercept variance is reflected in the hypothetical example in Figure 1 Panel B, wherein across clusters the intercepts are more variable at the extremes of $x_{\bullet j}$.[4] In sum, whereas a random slope of a purely level-1 predictor reflects *slope heterogeneity* across clusters (as illustrated in Figure 1 Panel A), a random slope of a purely level-2 predictor

[3] The conventional random-slope contextual effect model implies that the intercept variance follows a quadratic function of $x_{\bullet j}$, though it could in theory follow some other function (e.g., linear, cubic, etc.) which researchers could model instead. Here, however, I restrict focus to the structure implied by the conventional random-slope contextual effect model, as this model is commonly used and its issues are the primary focus of this dissertation.

[4] Here, and in Appendix A, I define the intercept as being conditional on the cluster mean of the level-1 predictor (see, e.g., Raudenbush, 1989). Alternatively, the intercept could be defined unconditionally (without conditioning on the cluster mean of the level-1 predictor) as $\gamma_{00} + u_{0j}$. The former definition is used to help explicate issues associated with random conflation (explained later), which arise regardless of which definition is used for the intercept.

reflects *intercept heteroscedasticity* (as shown in Figure 1 Panel B).

In the conventional random-slope contextual effect model, the random slope residual of x_{ij} (i.e., u_{1j}) thus simultaneously reflects a blend of *both* slope heterogeneity and intercept heteroscedasticity. That is, although the random slope variance in the conventional random-slope contextual effect model is $\text{var}(u_{1j}) = \tau_{11}$, the latter term τ_{11} also appears in the Equation (11) expression for the heteroscedastic intercept variance (i.e., $\text{var}(u_{0j} + u_{1j}x_{\bullet j} \mid x_{\bullet j}) = \tau_{00} + 2\tau_{01}x_{\bullet j} + \tau_{11}x_{\bullet j}^2$). Troublingly, in practice researchers interpret the variance τ_{11} from this conventional random-slope contextual effect model as representing purely slope heterogeneity (i.e., cluster-specific differences in slopes), and thus fail to recognize or note that it is conflated with intercept heteroscedasticity (e.g., Bliese & Britt, 2001; Deemer et al., 2017; Diez-Roux et al., 2000; Espelage et al., 2003; Fischer et al., 2004; Hoffman & Stawski, 2009; Kidwell et al., 1997; Lee, 2009; Lee & Bryk, 1989; Merlo et al., 2005; Poteat et al., 2007; Schempf & Kaufman, 2012; Titus, 2004).

To more precisely clarify the highly restrictive assumptions made by the conventional random-slope contextual effect model, here in Equation (12) I introduce separate notation for the level-specific random slope residuals (u_{wj} and u_{bj}), as well as for the level-specific fixed components (γ_w and γ_b).

$$y_{ij} = \gamma_{00} + \gamma_b x_{\bullet j} + \gamma_w(x_{ij} - x_{\bullet j}) + u_{bj}x_{\bullet j} + u_{wj}(x_{ij} - x_{\bullet j}) + u_{0j} + e_{ij}$$

$$e_{ij} \sim N(0, \sigma^2)$$

$$\begin{bmatrix} u_{0j} \\ u_{wj} \\ u_{bj} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \text{var}(u_{0j}) & & \\ \text{cov}(u_{0j}, u_{wj}) & \text{var}(u_{wj}) & \\ \text{cov}(u_{0j}, u_{bj}) & \text{cov}(u_{wj}, u_{bj}) & \text{var}(u_{bj}) \end{bmatrix} \right) \tag{12}$$

This representation in Equation (12) allows us to clarify that the conventional random-slope contextual effect model in Equation (9) makes the highly restrictive assumption that $u_{bj} = u_{wj}$, yielding random conflation.[5] In other words the random slope residual u_{1j} in Equation (9) implicitly reflects both u_{wj} and u_{bj}.

For further clarification, I can now represent this highly restrictive assumption of $u_{bj} = u_{wj}$ in the conventional random-slope contextual effect model of Equation (9) in terms of model parameters (noting that the residuals themselves are not parameters). Specifically, the conventional random slope contextual effect model of Equation (9) assumes:

(a) *equal variances* of u_{wj} and u_{bj} (i.e., $\text{var}(u_{bj}) = \text{var}(u_{wj})$)

(b) *perfect correlation* of u_{wj} and u_{bj} ($\text{corr}(u_{bj}, u_{wj}) = 1$).

In Appendix B, I prove that adding these constraints to the model in Equation (12) yields the conventional random-slope contextual effect model of Equation (9).

When these strict assumptions of equal variance and perfect correlation do not hold, as will be likely in practice, the estimated slope variance in the conventional random-slope contextual effect model is then an uninterpretable blend of its within-cluster component (slope heterogeneity) and its between-cluster component (intercept heteroscedasticity). Despite this, it is worth emphasizing that empirical applications exclusively interpret the estimated variance as slope heterogeneity without acknowledging its conflation with intercept heteroscedasticity (e.g., Bliese & Britt, 2001; Deemer et al., 2017; Diez-Roux et al., 2000; Espelage et al., 2003; Fischer et al., 2004; Hoffman & Stawski, 2009; Kidwell et al.,

[5] Note that the equality $u_{bj} = u_{wj}$ refers to equality of the true cluster-specific slope residuals, not the predictions (e.g., obtained via Empirical Bayes posterior modes) of those residuals.

1997; Lee, 2009; Lee & Bryk, 1989; Merlo et al., 2005; Poteat et al., 2007; Schempf & Kaufman, 2012; Titus, 2004). I will now provide concrete, graphical illustrations to communicate that this faulty but widespread interpretation of the conventional random-slope contextual effect model can lead to four kinds of erroneous inferences: 1) concluding there is no slope heterogeneity when it indeed exists; 2) concluding there is slope heterogeneity when it does not exist; 3) under- or overestimating the degree of slope heterogeneity; and 4) under- or overestimation of the standard errors for fixed components of slopes.

Chapter IV

Illustrating erroneous inferences that are consequences of using the conventional random-slope contextual effect model

Erroneous inference #1: Concluding there is no random slope variation when it exists

A key substantive goal of many empirical applications is to assess presence of cluster-specific differences in slopes (i.e., test the random slope variance τ_{11}), under the near-universally imposed assumption of intercept homoscedasticity, and a common fitted model for doing so is the conventional random-slope contextual effect model. I first illustrate that fitting this model *when the assumption of intercept homoscedasticity is indeed met* can nonetheless (for the specific reason described below) lead researchers to erroneously conclude that there is *no* slope heterogeneity when slope heterogeneity in fact exists. The reason is that, in this situation, the conflated residual u_{1j} from the conventional random-slope contextual effect model is weighted towards 0 because it implicitly reflects both the u_{wj} residuals (representing the pure slope heterogeneity that the researcher is substantively interested in) and the u_{bj} residuals (which are all 0 when there is intercept homoscedasticity). In turn, this creates less variability about 0 for the conflated residuals compared to the variability of the u_{wj}'s, leading to such underestimation of slope heterogeneity that the researcher can incorrectly infer there is none.

To illustrate, I generated a single sample[6] of data such that there was true across-

[6] In this section I present single sample graphical demonstrations for conceptual and concrete illustrative purposes, but in an upcoming section I follow up this demonstration with simulation results across many repeated samples, described subsequently in Table 2.

cluster slope variability (shown in Figure 2 Panel B) but no intercept heteroscedasticity (shown in Figure 2 Panel C) meaning that the intercept variance is constant across the range of $x_{\bullet j}$, as evident from the constant vertical spread of the intercepts (*y*-axis; obtained from the generating u_{bj}'s) across the range of $x_{\bullet j}$ (*x*-axis) in Figure 2 Panel C.[7] I then fit the conventional random-slope contextual effect model with *lmer* in R using restricted maximum likelihood (REML).[8] Figure 2 Panel A shows that in the fitted conventional random-slope contextual effect model the estimated slope variance $\hat{\tau}_{11}$ was near-zero (i.e., dramatically underestimated from the generating value of 2) and was non-significant (based on a χ^2 mixture likelihood ratio test [LRT] with the null reference distribution[9] being a 50:50 mixture

[7] In other words, I generated data from Equation 12 with variability in u_{wj} but no variability in u_{bj}, for 50 clusters of size 10. Generating parameters were: $\gamma_{00} = 1$, $\gamma_w = 3$, $\gamma_b = 1.5$, $\text{var}(u_{0j}) = 2$, $\text{var}(u_{wj}) = 2$, $\text{var}(u_{bj}) = 0$ and $\sigma^2 = 15$ (all other variance/ covariance components were 0). The level-1 variable x_{ij} was generated as the sum of a within-cluster component and a between-cluster component, each with variance 1. Data were generated from the observed cluster means and observation-specific deviations, consistent with standard MLM assumptions. Models were fit with *lmer* in R using restricted maximum likelihood (REML) estimation.

[8] REML was used to fit each model and obtain point estimates and standard errors because maximum likelihood (ML) provides biased estimation of random effect variances. ML was used separately, however, in computing the likelihood ratio test (LRT) statistics for the mixture chi-square tests of the random slope variances, as the derived null distributions for these tests assume ML estimation (Stram & Lee, 1994, 1995). Had I additionally used an LRT to test fixed components of slopes, I would again have needed to use ML, as REML is inappropriate to use when comparing models with different fixed effects (Snijders & Bosker, 2012).

[9] This null reference distribution was derived by Stram and Lee (1994, 1995) based on asymptotic theory. An alternative approach is to use an empirical null reference distribution. This can be done by repeatedly simulating data (with fixed sample size) from a model where the random slope variance is constrained to 0 and, for each sample, fitting models constraining versus not constraining the random slope variance, and then computing the likelihood ratio statistic between the model freely estimating the random slope variance and the model constraining it to 0. Then define the critical value as the value falling at the 95th percentile of the empirical null distribution and check whether the observed LRT statistic exceeds that critical value (see, e.g., *RLRsim* in R; Scheipl & Bolker, 2013).

of $\chi^2_{df=1}$ and $\chi^2_{df=2}$;[10] Stram & Lee, 1994, 1995). This apparent lack of slope heterogeneity is reflected in the plot in Figure 2 Panel A, as all of the estimated cluster-specific regression lines of y_{ij} on x_{ij} (obtained from empirical Bayes predictions/ posterior modes via the *ranef* function in R) are nearly parallel. When fitting the conventional random-slope contextual effect model, then, an applied researcher would erroneously substantively conclude there is no cluster-specific variability in slopes, despite the fact that there is indeed across-cluster slope heterogeneity, as apparent from the nonparallel cluster-specific regression lines of y_{ij} on $x_{ij} - x_{\bullet j}$ in Figure 2 Panel B (obtained from the generating u_{wj}'s). If the conventional random-slope contextual effect model were able to recover the cluster-specific variation in slopes, the cluster-specific regression lines in Panel A would have had the same slopes as those from Panel B, noting that plotting regression lines by x_{ij} (Panel A) vs. $x_{ij} - x_{\bullet j}$ (Panel B) would impact only the location of the lines but not the slopes themselves, as $x_{\bullet j}$ is constant with respect to cluster.

Erroneous inference #2: Concluding that there is slope heterogeneity when it does not exist

Next I illustrate another erroneous inference that researchers can make when fitting the conventional random-slope contextual effect model: concluding that there is slope heterogeneity when there is none. For this demonstration, I generated a single sample with a fixed slope (i.e., all clusters have a u_{wj} of 0, shown in Figure 3 Panel B) but with intercept

[10] The two values for the chi-square degrees of freedom in the 50:50 mixture are the number of rows/columns in the random effect covariances matrix for each of the two models under comparison (Stram & Lee, 1994). The conventional random slope contextual effect model has a 2X2 random effect covariance matrix (hence *df*=2), whereas the model constraining the random slope variance to 0 has only a single random effect (intercept) variance (hence *df*=1).

heteroscedasticity (i.e., variability in u_{bj}, shown in Figure 3 Panel C).[11] Because the slope residuals u_{1j} in the conventional random-slope contextual effect model are conflated, reflecting both the actual u_{wj} and u_{bj} residuals, the conflated residuals are in this case weighted away from 0 compared to the u_{wj}'s. This creates more variability about 0 for the conflated residuals compared to the variability of the u_{wj}'s, leading to overestimating the slope heterogeneity (which here is actually nonexistent) and incorrectly inferring that there are across-cluster difference in slopes.

Figure 3 Panel A shows the conventional random-slope contextual effect model results, wherein there is apparent slope heterogeneity, as the $\text{var}(u_{1j})$ estimate is significant and the cluster-specific regression lines of y_{ij} on x_{ij} (obtained from empirical Bayes predictions/ posterior modes) are markedly nonparallel. Thus, when fitting the conventional random-slope contextual effect model to this sample dataset, an applied researcher would make the erroneous substantive conclusion that there is across-cluster variability in slopes, when in fact there is only intercept heteroscedasticity.

Erroneous inference #3: Under- or overestimating the degree of slope heterogeneity

From the two demonstrations thus far, I have shown how fitting a conventional random-slope contextual effect model in the presence of intercept homoscedasticity can (for the specific aforementioned theoretical reasons) lead one to incorrectly infer there is no slope heterogeneity (which will correspond with inflated type II error in later full-scale simulations) whereas doing so in the presence of intercept heteroscedasticity can lead one to

[11] Generating parameters were the same as described in the footnote 7, with the exception that here $\text{var}(u_{wj}) = 0$, $\text{var}(u_{bj}) = 8$, and $\sigma^2 = 10$.

incorrectly infer there is slope heterogeneity (which will correspond with inflated type I error in later full-scale simulations). These demonstration situations by definition involved either u_{wj} or u_{bj} having no variability. I now explain and illustrate how erroneous conclusions about the degree of slope heterogeneity still arise under more general conditions wherein there can be variability in *both* of the level-specific residuals (u_{wj} and u_{bj}) that contribute to the conflated residual u_{1j} in the conventional random-slope contextual effect model. Specifically, here three such conditions are considered in which: (a) the variance of u_{wj} is equal to that of u_{bj}, (b) the variance of u_{wj} is greater than that of u_{bj}, and (c) the variance of u_{wj} is less than that of u_{bj}. For each of these conditions, I vary the correlation of the two residuals (u_{wj} and u_{bj}) across its possible range. For each of these conditions, I illustrate how, and then theoretically explain why, the slope variance τ_{11} from the conventional random slope contextual effect model can either *overestimate* or *underestimate* the true slope heterogeneity depending upon both the magnitude of the variance of u_{bj} and the degree of correlation between u_{bj} and u_{wj}.

Condition 1: $\mathrm{var}(u_{wj}) = \mathrm{var}(u_{bj})$. Recall that the conventional random-slope contextual effect model makes the assumptions that $\mathrm{var}(u_{wj}) = \mathrm{var}(u_{bj})$ and $\mathrm{corr}(u_{bj}, u_{wj}) = 1$. Here I generate data such that this first assumption is met in that both variances are equal to 2, but manipulate the correlation between the two residuals across the parameter space (either: -1, -.95, -.9 -.75, -.5, -.25, 0, .25, .5, .75, .9, .95, or 1).[12]

Figure 4 Panel A column 1 shows the percent relative bias in estimating the slope

[12] The other generating values are the same as in footnote 5.

variance, defined as the difference in the average estimated slope variance across 5000 samples, $\overline{\text{var}}(u_{1j})$, and the true across-cluster slope variance, $\text{var}(u_{wj})$, divided by $\text{var}(u_{wj})$, and multiplied by 100 (i.e., $100\times(\overline{\text{var}}(u_{1j})-\text{var}(u_{wj}))/\text{var}(u_{wj})))$ at each $\text{corr}(u_{wj},u_{bj})$. The two horizontal lines denote where absolute relative bias is 10% or -10%; values outside of this range conventionally indicate meaningfully substantial bias.[13]

Note first that, from Figure 4 Panel A column 1, there is no bias in the estimation of the random slope variance τ_{11} when the two strict assumptions of the conventional random-slope contextual effect model are met, as the thin (red) line is at roughly 0 when $\text{corr}(u_{bj},u_{wj})=1$. However, this is the only condition across the range of the correlation space in which this is true. The further the $\text{corr}(u_{bj},u_{wj})$ is from 1, the more downward bias there is, meaning that the conventional random-slope contextual effect model is *underestimating* the true degree of slope heterogeneity. As is clear from the plot, this parameter bias can be quite large, for instance, approaching -50% relative bias as $\text{corr}(u_{bj},u_{wj})$ decreases.

To explain why this occurs, consider first the extreme of $\text{corr}(u_{bj},u_{wj})=-1$. In this situation, a cluster with a large u_{wj} will have a small u_{bj}, and vice versa. For instance, if a cluster has a u_{wj} of 1, the u_{bj} would be -1. The conflated residual from the conventional random-slope contextual effect model, u_{1j}, is reflecting both of these. Thus, in comparison to

[13] Though a rare occurrence, when models failed to converge or provided improper solutions, these samples were excluded. In an upcoming section, I consider results for each of five different models, and excluded samples in which any of the five models did not converge or provided improper solutions. For all simulations and all conditions in this paper, at least 96% of samples were retained.

u_{wj}, the u_{1j} is pulled sharply towards 0, and hence u_{1j} has less variability about 0 than u_{wj}, leading to underestimation. As the correlation gets larger, this effect is less dramatic; for instance, at a correlation of -.5, if a cluster has a u_{wj} of 1, the u_{bj} might be -0.5, and thus in comparison to the situation with a correlation of -1, the conflated residual is pulled towards 0 less sharply. However, only when the correlation is exactly 1 does the bias disappear; in this case, a cluster with a u_{wj} of 1 would also have a u_{bj} of 1, and thus the conflated residual u_{1j} would be the same as u_{wj}.

Condition 2: $\text{var}(u_{wj}) > \text{var}(u_{bj})$. Next I consider the case wherein the equal variance assumption ($\text{var}(u_{wj}) = \text{var}(u_{bj})$) is not met because the degree of slope heterogeneity is greater than the degree of intercept heteroscedasticity, i.e., $\text{var}(u_{wj}) > \text{var}(u_{bj})$. Here I generate data as in *Condition 1* above (changing only $\text{var}(u_{bj})$ to 1) and similarly fit the conventional random-slope contextual effect model to 5000 samples to determine the percent relative bias across the range of $\text{corr}(u_{bj}, u_{wj})$. The results are presented in Figure 4 Panel A column 2. Though here the pattern is similar to that from *Condition 1* in that the bias is mitigated as the correlation increases, unlike *Condition 1*, the conventional random-slope contextual effect model *always substantially underestimates* the true slope heterogeneity.

To explain why this occurs even when $\text{corr}(u_{bj}, u_{wj}) = 1$, consider that, in this case, if a cluster has a u_{wj} of 1, it would similarly have a positive u_{bj} but of smaller magnitude, i.e., 0.5. Because the conflated residual from the conventional random-slope contextual effect model, u_{1j}, reflects both of these, it would be somewhere between the two. Across clusters, then, the conflated residuals u_{1j} will be closer to 0 than the u_{wj}'s, causing downward bias in the estimated slope variance. A key point illustrated here is that, whenever $\mathrm{var}(u_{bj})$ is not exactly equal to $\mathrm{var}(u_{wj})$, the assumptions of the conventional random-slope contextual effect model will never be met, regardless of the correlation between the two residuals, and thus the conflated random slope will always be some uninterpretable blend of the two components.

Condition 3: $\mathrm{var}(u_{wj}) < \mathrm{var}(u_{bj})$. Lastly I consider the case in which again the equal variance assumption of the conventional random-slope contextual effect model is not upheld, in this case because there is a larger degree of intercept heteroscedasticity than slope heterogeneity, i.e., $\mathrm{var}(u_{wj}) < \mathrm{var}(u_{bj})$. Here I generate data as in *Condition 1*, but with $\mathrm{var}(u_{bj})$ set to 8. Also as in *Conditions 1* and *2*, I manipulated $\mathrm{corr}(u_{bj}, u_{wj})$ and computed the percent relative bias across the range of $\mathrm{corr}(u_{bj}, u_{wj})$ with 5000 samples. The results are presented in Figure 4 Panel A column 3. Similar to *Conditions 1* and *2*, the across-sample average slope variance estimate increases as $\mathrm{corr}(u_{bj}, u_{wj})$ increases. However, unlike *Conditions 1* and *2*, the true slope heterogeneity is *underestimated* at lower values of $\mathrm{corr}(u_{bj}, u_{wj})$ but *overestimated* at higher values.

I wish to emphasize this possibility of overestimation, as previous authors have stated that the slope variance of x_{ij} will be smaller than that of $x_{ij} - x_{\bullet j}$ (e.g., Enders & Tofighi, 2007; Raudenbush & Bryk, 2002), though this is clearly not always the case, as shown in Figure 4 Panel A column 3. As for why this overestimation is also possible, consider the extreme of $\mathrm{corr}(u_{bj}, u_{wj}) = 1$. In this condition, a cluster with a u_{wj} of 1 would have a u_{bj} of 2. The conflated residual u_{1j} in the conventional random-slope contextual effect model then would be in between these two values, and thus relative to the u_{wj}, there is more variability of u_{1j} about 0, leading to a conflated slope variance estimate that is greater than $\text{var}(u_{wj})$. Thus, anytime there is a large amount of intercept heteroscedasticity relative to slope heterogeneity (i.e., $\text{var}(u_{wj}) < \text{var}(u_{bj})$), there is risk of *overestimating* the degree of slope heterogeneity when u_{bj} and u_{wj} are positively correlated.

Summary. The key point to these demonstrations in Figure 1 Panel A is that the slope variance from the conventional random-slope contextual effect model can either underestimate or overestimate the degree of actual slope heterogeneity, as it instead reflects variance in an uninterpretable blend of cluster-specific random slope components (u_{wj}'s) and heteroscedastic intercept components (u_{bj}'s). Researchers can be assured that this does not happen only when the two components are perfectly correlated and have equal variances.

Erroneous inference #4: Under- or overestimating standard errors of fixed effects

Thus far I have considered the impact of random conflation on the testing and estimation of the random slope variance itself (e.g., presence and degree of cluster-specific differences in slopes). It is important for researchers to realize, however, that random conflation can adversely affect the fixed portion of the model as well, in that it can yield inappropriate standard errors for fixed components of slopes. Though it is well known in the MLM literature that misspecification of the random effect structure can yield inappropriate standard errors for fixed effects (e.g., Raudenbush & Bryk, 2002; Snijders & Bosker, 2012), to my knowledge, no one has demonstrated the adverse impact of random conflation specifically. Hence, for all conditions described above in the Erroneous Inference #3 section, I obtained the average standard errors across 5000 repeated samples for both the within-cluster fixed effects (the $\hat{\gamma}_{10}$'s) and the contextual effects (the $\hat{\gamma}_{01}$'s) and compared them to the actual across-5000-repeated-samples standard deviation of $\hat{\gamma}_{10}$ and $\hat{\gamma}_{01}$, respectively, by computing percent relative bias. Percent relative bias is here defined as the difference in the average standard error and the across-sample standard deviation divided by the across-sample standard deviation, multiplied by 100.

Bias in the standard error for the within-cluster fixed effect. Figure 4 Panel B shows the percent relative bias for the standard error of $\hat{\gamma}_{10}$ from the conventional random-slope contextual effect model across the range of $\text{corr}(u_{bj}, u_{wj})$ for each of the three conditions: $\text{var}(u_{wj}) = \text{var}(u_{bj})$, $\text{var}(u_{wj}) > \text{var}(u_{bj})$, and $\text{var}(u_{wj}) < \text{var}(u_{bj})$. In short, the pattern of results in Figure 4 Panel B mirrors exactly the pattern found for the random slope variance estimates in Figure 4 Panel A—whenever the random slope variance is *underestimated*, the standard errors for $\hat{\gamma}_{10}$ are *too small*, and whenever the random slope variance is *overestimated*, the standard errors for $\hat{\gamma}_{10}$ are *too large*. This standard error bias is often substantial and can be as severe as -20%. In empirical practice, such standard error bias in the conventional random-slope contextual effect model would naturally adversely impact confidence interval estimation and inferential testing of the fixed effects, leading to the potential for misleading substantive conclusions. Only when the strict assumptions of $\text{var}(u_{wj}) = \text{var}(u_{bj})$ and $\text{corr}(u_{bj}, u_{wj}) = 1$ are met can researchers be assured that the standard errors are accurate (as shown at the right-most point in Figure 4 Panel B column 1).[14]

[14] The bias in the point estimates of fixed effect was negligible, as it was for all the other models fit in the upcoming simulation. This finding is expected based on previous analytic and simulation findings that the fixed effects are consistent and tend also to be unbiased in spite of random effect misspecification (e.g., Maas & Hox, 2004; Verbeke & Lessafre, 1997). Hence I focus specifically here on standard error estimation.

Bias in the standard error for the contextual effect. Similarly, Figure 4 Panel C depicts the standard error percent relative bias for the contextual effect, $\hat{\gamma}_{01}$, across the range of $\text{corr}(u_{bj}, u_{wj})$ for each of the three previously described conditions. In the equal variance condition (column 1), we can see that the standard errors are accurate when the conventional random-slope contextual effect model assumptions are exactly met. This is, however, generally not the case. Since $\text{var}(u_{bj})$ is (implicitly) estimated as the random slope variance of x_{ij} (because the u_{1j}'s simultaneously reflect both the u_{wj}'s and the u_{bj}'s), and given that the u_{1j}'s are generally weighted away from the u_{bj}'s towards 0, the random slope variance of x_{ij} generally underestimates $\text{var}(u_{bj})$, leading to downward bias in the standard errors of $\hat{\gamma}_{01}$. Though the bias is mitigated for each condition as $\text{corr}(u_{bj}, u_{wj})$ increases, the standard errors are almost always underestimated to an extreme degree. For instance, as shown in Figure 4 Panel C column 3, when there is a large degree of heteroscedasticity, failing to model this appropriately yields percent relative bias for the contextual effect standard error that can be worse than -40%. Troublingly, such standard error bias could lead applied researchers to make erroneous inferences about the substantively important contextual effect.

Summary. Though previous sections' demonstrations concerned estimation and testing of the random slope variance itself, researchers need to understand that random conflation can have adverse impacts elsewhere in the model as well. As I have shown in this section, standard errors of substantively important fixed effects can be highly inaccurate under such random conflation. Researchers fitting the conventional random-slope contextual effect model (or uncentered random slope model) can be assured of avoiding such standard error bias only if the two random components u_{wj} and u_{bj} are perfectly correlated and have exactly equal variances.

Chapter V

General taxonomy of slope conflation in multilevel models

Though discussed thus far in the context of the conventional random-slope contextual effect model, the terms *fixed conflation* and *random conflation* can be used more generally to describe and classify any random-slope MLM that contains level-1 variables. Here I do so to provide a general taxonomy of slope conflation for MLMs. Recall that anytime a MLM contains x_{ij}, one can re-express the model by separating x_{ij} into its pure level-specific parts, $x_{ij} - x_{\bullet j}$ (often called the group-mean-centered predictor) and $x_{\bullet j}$ (the group mean). By doing this, one can always express the fixed components of a level-1 variable in the form $\gamma_w(x_{ij} - x_{\bullet j}) + \gamma_b x_{\bullet j}$ and the random components of a level-1 variable in the form $u_{wj}(x_{ij} - x_{\bullet j}) + u_{bj}x_{\bullet j}$. I propose the following taxonomy:

- The MLM is *fully conflated* if it constrains $\gamma_b = \gamma_w$ and $u_{bj} = u_{wj}$.
- The MLM is *partially conflated* if it constrains $\gamma_b = \gamma_w$ or $u_{bj} = u_{wj}$, but not both. A partially conflated model can be further characterized as either having fixed conflation or random conflation.[15]
- The MLM is *unconflated* if neither of these constraints are made.

For instance, the conventional random-slope contextual effect model of Equation (9) constrains $u_{bj} = u_{wj}$ but does not constrain $\gamma_b = \gamma_w$ and is thus a partially conflated model.

[15] Note that a partially conflated model that conflates only the $u's$ not the $\gamma's$ is much more common in practice than a partially conflated model that conflates only the $\gamma's$ not the $u's$.

Methods for unconflating random slopes

I begin by summarizing the different methods by which researchers can ensure a random slope model is unconflated; subsequently, I provide detailed explanations of these methods. For random-slope contextual effect MLMs (or uncentered MLMs which contain a random slope of x_{ij}), random conflation can be avoided by including a random component of $x_{\bullet j}$. This is almost never done in empirical applications because intercept heteroscedasticity is rarely substantively theorized (e.g., Goldstein, 2011; Rights & Sterba, 2016; Snijders & Bosker, 2012). On the other hand, for random-slope group-mean-centered MLMs (which contain a random slope of $x_{ij} - x_{\bullet j}$), random conflation is already *inherently avoided* because there is no implicit equality constraint of $u_{wj} = u_{bj}$ (as will be shown below), hence unconflation is assured whether or not a random component of $x_{\bullet j}$ is included. These methods for unconflating random slopes are summarized in Table 1. For each possible random slope MLM specification, Column 4 of Table 1 clarifies whether fixed and/or random components are conflated. The first row of Table 1 pertains to uncentered MLMs, the next two rows pertain to contextual effect MLMs, and the last three rows pertain to group-mean-centered MLMs. Column 4 of Table 1 highlights that, for uncentered and contextual effect MLMs (first three rows) the modeling of between-cluster variance via fixed and/or random slope components for $x_{\bullet j}$ is a mechanism for unconflation, whereas in the group-mean-centered MLMs (last three rows), it is a separate decision in that it does not impact whether or not there is random conflation, as conflation is inherently avoided. Table 1 also characterizes MLM specifications by whether they have been commonly used (Column 5) and whether they have been previously recommended (Column 6). Columns 5 and 6 of Table

1 highlight the noteworthy and concerning fact that the commonly used and previously recommended MLMs include those that are random conflated, such as the conventional random-slope contextual effect model, featured in row 2 of Table 1. Column 7 of Table 1 highlights that only the unconflated MLM specifications are currently recommended here, and these include specifications not yet recognized or used for this purpose in practice (i.e., row 3). The alternative unconflated MLM specifications are explained in detail next.

To avoid random conflation in models that contain x_{ij} as a predictor, one must expand the Equation (9) MLM to explicitly model intercept heteroscedasticity by including a random component of a slope for $x_{\bullet j}$. This is done in the specification in Table 1 row 3, which is here termed the *unconflated random-slope contextual effect model*. This model is given as:

$$
\begin{gathered}
y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{1j}x_{ij} + \gamma_{01}x_{\bullet j} + u_{2j}x_{\bullet j} + u_{0j} + e_{ij} \\
e_{ij} \sim N(0,\sigma^2) \\
\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ \tau_{20} & \tau_{21} & \tau_{22} \end{bmatrix} \right)
\end{gathered} \tag{13}
$$

with u_{2j} denoting the residual used in modeling intercept heteroscedasticity. The disaggregation of both the fixed and random components is apparent when substituting x_{ij} with $(x_{ij} - x_{\bullet j}) + x_{\bullet j}$:

$$
\begin{aligned}
y_{ij} &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\bullet j} + x_{\bullet j}) + u_{1j}(x_{ij} - x_{\bullet j} + x_{\bullet j}) + \gamma_{01}x_{\bullet j} + u_2 x_{\bullet j} + u_{0j} + e_{ij} \\
&= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\bullet j}) + u_{1j}(x_{ij} - x_{\bullet j}) + (\gamma_{10} + \gamma_{01})x_{\bullet j} + (u_{1j} + u_2)x_{\bullet j} + u_{0j} + e_{ij}
\end{aligned} \tag{14}
$$

The fixed component is disaggregated because $\gamma_w \neq \gamma_b$ (i.e., $\gamma_{10} \neq (\gamma_{10} + \gamma_{01})$), but now additionally the random component is disaggregated because $u_{wj} \neq u_{bj}$ (i.e., $u_{1j} \neq (u_{1j} + u_{2j})$).

This model is virtually never applied in practice because researchers have not recognized the need to incorporate intercept heteroscedasticity for the pragmatic purpose of unconflating the random component, and (according to Goldstein [2011], Rights & Sterba [2016], Snijders & Berkhof [2008] and Snijders & Bosker [2012]) researchers rarely if ever incorporate intercept heteroscedasticity for other, purely substantive reasons.

I next consider group-mean-centered models that explicitly include as a predictor $x_{ij} - x_{\bullet j}$ rather than x_{ij}. These are given in rows 4-6 of Table 1. Thus far, I have not discussed group-mean-centered models at length. The reason is that, when group-mean-centering, *neither fixed nor random conflation is possible*. In other words, the issues I have discussed thus far are irrelevant for group-mean-centered models as they are inherently unconflated. To illustrate, I first present the most parsimonious such unconflated model which was termed a *random-slope group-mean-centered MLM* in Table 1 row 4:

$$y_{ij} = \gamma_{00} + \gamma_{10}(x_{ij} - x_{\bullet j}) + u_{1j}(x_{ij} - x_{\bullet j}) + u_{0j} + e_{ij}$$

$$e_{ij} \sim N(0, \sigma^2)$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix} \right) \tag{15}$$

To determine whether or not this model is conflated, I need not do the substitution done for the uncentered or contextual effect models. When group-mean-centering, variables are already separated into pure level-specific portions. Here, the fixed component is not conflated because $\gamma_w \neq \gamma_b$ (i.e., $\gamma_{10} \neq 0$), and the random component is not conflated because $u_{wj} \neq u_{bj}$ (i.e., $\gamma_{10} \neq 0$).

The above unconflated group-mean-centered model, however, is useful only for estimating the within-cluster effect of x_{ij} and not the between-cluster effect. Though not necessary for the purposes of unconflation, if a researcher wanted to additionally model a between-cluster fixed effect they could add a fixed slope of $x_{\cdot j}$. This is done in Table 1 row 5, there termed a *random-slope group-mean-centered MLM with fixed between effect*:

$$y_{ij} = \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j}) + u_{1j}(x_{ij} - x_{\cdot j}) + \gamma_{01}x_{\cdot j} + u_{0j} + e_{ij}$$

$$e_{ij} \sim N(0, \sigma^2)$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix} \right) \tag{16}$$

Lastly, though again not necessary for the purpose of unconflation, this unconflated model can be further expanded to account for intercept heteroscedasticity by adding a random component to the slope of $x_{\cdot j}$. This is shown in Table 1 row 6, there termed a *heteroscedastic random-slope group-mean-centered MLM with between effect*:

$$y_{ij} = \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j}) + u_{1j}(x_{ij} - x_{\cdot j}) + \gamma_{10}x_{\cdot j} + u_{2j}x_{\cdot j} + u_{0j} + e_{ij}$$

$$e_{ij} \sim N(0, \sigma^2)$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ \tau_{20} & \tau_{21} & \tau_{22} \end{bmatrix} \right) \tag{17}$$

Note that this model is statistically equivalent to the unconflated random-slope contextual effect model in Equation (13), as the two are merely different parameterizations (as proven in Appendix C). All other models in Table 1 are nested within these two. The uncentered random slope model (row 1) and the conventional random-slope contextual effect model (row 2) are simpler in that they implicitly constrain level-specific effects to equality (i.e., conflate); similarly, the group-mean-centered models in rows 4 and 5 are simpler in that they implicitly constrain between-cluster effects to 0.

In summary, for models that include a random slope of x_{ij}, whether or not the model includes fixed and/or random slope components for $x_{\bullet j}$ determines whether or not there is fixed conflation and/or random conflation. In contrast, for group-mean-centered models, both fixed conflation and random conflation are inherently avoided regardless of the inclusion of fixed and/or random slope components for $x_{\bullet j}$. I next present simulation results comparing the performance of these models and then provide recommendations for practice.

Chapter VI

Comparing the performance of the conventional random-slope contextual effect MLM to the unconflated MLMs

Earlier, I explained why the conventional random-slope contextual effect model (Equation 9; Table 1 row 2) can yield low power for testing across-cluster slope heterogeneity (Erroneous Inference #1), high Type I error for testing across-cluster slope heterogeneity (Erroneous Inference #2), can either under- or overestimate the true degree of across-cluster slope heterogeneity (Erroneous Inference #3), and can yield inaccurate standard errors for fixed effects (Erroneous Inference #4). Here I show how the unconflated models (Equations 13, 15, 16, 17; Table 1 rows 3-6) provide markedly better performance in all four of these domains. To restrict focus on the implications of random conflation, I did not additionally fit the uncentered random-slope model (Equation 7; Table 1 row 1) that also conflates the fixed component; nonetheless note that it is susceptible to the same four issues as the conventional random-slope contextual effect model.

Unconflated models provide better power for testing across-cluster slope heterogeneity. Using the generating conditions described above in the *Erroneous Inference #1* section, I generated 5000 samples and fit each of the models in Table 1 rows 2-6. Across the 5000 samples, I computed the proportion of samples wherein the random slope variance of the level-1 predictor (x_{ij} for the contextual effect models and $x_{ij} - x_{\bullet j}$ for the group-mean-centered models) was significant based on a χ^2 mixture LRT with $\alpha = .05$.[16] Since there

[16] The null reference distribution for the unconflated random-slope contextual effect model and the heteroscedastic random-slope group-mean-centered model was a 50:50 mixture of

exists slope heterogeneity in this condition ($\text{var}(u_{wj}) = 2$), this proportion denotes the power to reject the false null hypothesis of no slope heterogeneity. As is clear from Table 2 column 1, the widely used conventional random-slope contextual effect model yields much lower power than the unconflated models (which include the proposed unconflated random-slope contextual effect model as well as any random-slope group-mean-centered model). Because these unconflated models explicitly disaggregate the u_{wj} and u_{bj}, unlike the conventional random-slope contextual effect model, they do not encounter the same problem as described in the above *Erroneous Inference #1* section.

Unconflated models provide better type I error for testing across-cluster slope heterogeneity. Using the generating conditions described above in the previous *Erroneous Inference #2* section, I again generated 5000 samples, fit each of the models in Table 1 rows 2-6, and for each computed the proportion of samples in which the random slope variance of the level-1 predictor was significant. Since there is no slope heterogeneity in this generating condition ($\text{var}(u_{wj}) = 0$), this proportion denotes the Type I error rate. As is clear from Table 2 column 2, the widely used conventional random-slope contextual effect model yields a large Type I error rate (nearly three times the nominal rate of .05) whereas the unconflated models (which include the proposed unconflated random-slope contextual effect model as well as any group-mean-centered model) yield a rate close to the nominal level. This is because unconflated models are not susceptible to the issue described in the earlier *Erroneous Inference #2* section.

Unconflated models provide less bias in estimating the degree of across-cluster slope

$\chi^2_{df=2}$ and $\chi^2_{df=3}$; for all other models, it was a 50:50 mixture of $\chi^2_{df=1}$ and $\chi^2_{df=2}$ (Stram & Lee, 1994, 1995).

heterogeneity. Using the generating conditions described above in the *Erroneous Inference #3* section, I again generated 5000 samples, fit each of the models in Table 1 rows 2-6, and for each computed the across-sample average random slope variance of the level-1 predictor. Figure 5 Panel A shows that, regardless of the degree of intercept heteroscedasticity (i.e., magnitude of $\text{var}(u_{bj})$) and regardless of the correlation of u_{wj} and u_{bj}, all of the unconflated models yield negligible bias in estimating the degree of slope heterogeneity. I note one small caveat in that, at the boundary of the correlation space (i.e., $\text{corr}(u_{wj}, u_{bj}) = -1$ or $\text{corr}(u_{wj}, u_{bj}) = 1$), there is very slight upward bias for the models that include a random component for $x_{\bullet j}$. This is not surprising given that, simply because of the boundary condition, the estimated random effect covariance matrix for these will slightly overestimate the covariance between u_{bj} and u_{wj} when $\text{corr}(u_{wj}, u_{bj}) = -1$ and slightly underestimate the covariance when $\text{corr}(u_{wj}, u_{bj}) = 1$. Nonetheless, the percent relative bias at the boundary is still fairly small (at most, 5%), particularly in comparison to the bias across the correlation space from the widely used conventional random-slope contextual effect model (which reaches -20%).

Unconflated models provide more accurate standard errors for fixed components of slopes. Using the generating conditions described above in the earlier *Erroneous Inference #4* section, I again generated 5000 samples, fit each of the models in Table 1 rows 2-6, and for each computed the across-sample average standard error for both the fixed component of the slope of the level-1 predictor and that of the level-2 predictor, comparing these values to the actual across-sample standard deviation in terms of percent relative bias. For the fixed component of the slope of the level-1 predictor, for all unconflated models (which include the proposed unconflated random-slope contextual effect model as well as any group-mean-centered model) the standard errors are accurate in Figure 5 Panel B regardless of the degree of intercept heteroscedasticity or the correlation of u_{wj} and u_{bj}.

For the fixed component of the level-2 predictor, the unconflated models that contain a random component for $x_{\bullet j}$ (such as the proposed unconflated random-slope contextual effect model) similarly provided accurate standard errors in Figure 5 Panel C (with a slight exception again being at the boundary of the correlation space). In contrast, it is important to note that in Figure 5 Panel C the random-slope group-mean-centered model that includes a fixed slope of $x_{\bullet j}$ without a random component (as in Equation (16); Table 1 row 5), provided standard errors that were too small, and the magnitude of this bias was greater when the degree of heteroscedasticity $(\text{var}(u_{bj}))$ was greater. Thus, this latter model allows for accurate inference about level-2 fixed effects only when the assumption of intercept homoscedasticity is upheld.

Summary of simulation results. In summary, when fitting the conventional random-slope contextual effect models, failing to include a random component of $x_{\bullet j}$ induces random conflation leading to Erroneous Inferences #1-4 described above; these issues can be avoided, however, by fitting the unconflated models. Specifically, fitting the proposed unconflated random-slope contextual effect model (which includes a random component to the slope of $x_{\bullet j}$) or fitting group-mean-centered models (regardless of whether $x_{\bullet j}$ is included, has a fixed slope, or has a random slope) will provide unbiased estimates, adequate power, acceptable type I error, and appropriate standard errors for all terms associated with the level-1 predictor, $x_{ij} - x_{\bullet j}$. Results from the group-mean-centered models reflect the key fact that $x_{ij} - x_{\bullet j}$ and $x_{\bullet j}$ are orthogonal, and hence misspecifying the portion of the model dealing with $x_{\bullet j}$ does not adversely impact the estimation of components associated with $x_{ij} - x_{\bullet j}$, such as τ_{11}. However, misspecifying the portion of the model dealing with $x_{\bullet j}$ in group-mean-centered models, such as failing to include a random component of $x_{\bullet j}$ under heteroscedasticity, can adversely impact inference about parameters associated with level-2 variables; for instance, leading to standard errors of the fixed component of $x_{\bullet j}$ that are too small, as in Figure 5 Panel C.

Chapter VII

Recommendations for practice: Choosing among the unconflated MLMs

In light of these simulation results, I do not recommend conflated models be used in practice, as noted in Table 1 Column 7 (in the Discussion section, however, I discuss the potential for testing conflated vs. unconflated models in practice). Unfortunately, both of the conflated models I have discussed are commonly used in practice, as noted in Table 1 Column 5 (rows 1 and 2). In fact, as noted in in Table 1 Column 6, methodologists have even specifically recommended that the conventional random-slope contextual effect model be used to disaggregate level-specific effects instead of using group-mean-centering, claiming that the former is less complicated and more interpretable (discussed further below)—however, importantly they failed to note the random conflation inherent to this model (e.g., Hox, 2010; Kreft, de Leeuw, & Aiken, 1995; Snijders & Bosker, 2012). It is my hope that this work will help discourage the use of such random-conflated models. As mentioned earlier, other authors have pointed out the fact that the conventional random-slope contextual effect model is random conflated—Raudenbush (1989) called it a "hidden covariance" between the intercept and slope—and recommended group-mean-centered models be used in practice (Raudenbush, 1989; Wang & Maxwell, 2015; Hofmann & Gavin, 1998). Here I have extended their work by delineating the various consequences of random conflation, as well as providing multiple options to avoid it.

Regarding which of the unconflated specifications to use, the decision between the unconflated random-slope contextual effect model and the heteroscedastic group-mean-centered model is primarily a matter of interpretational preference, as the two are reparameterizations of each other. One may prefer to interpret the fixed component of the slope of $x_{\bullet j}$ as the contextual effect (in the former model), or instead may wish to interpret it as the fixed between effect (in the latter model). Likewise one may prefer to interpret the random component of the slope of $x_{\bullet j}$ (i.e., the cluster-specific residual u_{2j}) as the cluster-specific difference in u_{wj} and u_{bj}, akin to a cluster-specific contextual effect residual (in the former model), or instead may wish to interpret it as the heteroscedastic intercept component, or u_{bj} (in the latter model).

Some authors have argued that the interpretation of contextual effect models in general is more straightforward than group-mean-centered models, and that testing the existence of a contextual effect is easier as it involves simply testing if the contextual effect is 0 (e.g., Algina & Swaminathan, 2011; Hox, 2010; Kreft, de Leeuw, & Aiken, 1995). I feel, however, that parameter interpretation in group-mean-centered models is no more complicated than in contextual effect models as the presence of a contextual effect can also easily be detected by testing the equality of the fixed within- and between-cluster effects (e.g., Enders & Tofighi, 2007; Wang & Maxwell, 2015).

A potential advantage of group-mean-centered models, however, is that researchers can specify more parsimonious models without conflating level-specific effects. For instance, if a researcher is solely interested in effects at level-1, then there is no need to include $x_{\bullet j}$ as a predictor in the model, and they can thus fit the group-mean-centered model that excludes $x_{\bullet j}$ (Equation (15); Table 1 row 4). If a researcher is interested in both level-1 and level-2 fixed effects and is comfortable assuming intercept homoscedasticity, they can fit the group-mean-centered model that includes only a fixed slope of $x_{\bullet j}$ with no random component (Equation (16); Table 1 row 5). It may be pragmatic, however, to test this assumption of homoscedasticity, as simulation results indicate that the standard errors for the fixed slope of $x_{\bullet j}$ would be inappropriate if there is indeed intercept heteroscedasticity.

Chapter VIII

Empirical example

To concretely illustrate the differences in interpretation between fully conflated, partially conflated, and unconflated MLM specifications, here I predict math scores using a dataset described and analyzed in a popular multilevel modeling textbook (Kreft & de Leeuw, 1998). The dataset consists of 519 students (level-1) nested with 23 schools (level-2), yielding an average cluster size of approximately 23. Between-school variation accounted for 36% of the total variation in math scores (*ICC* = .36). Kreft and de Leeuw (1998) predicted math scores from parents' highest level of education with the fully conflated uncentered random-slope MLM of Equation (7) (Table 1 row 1). Here I show how results can be made more informative by unconflating both the fixed and the random components of parent education.

I first fit the (fully conflated) uncentered random slope MLM in which uncentered parent education predicts math scores (i.e., Equation (7) , Table 1 Row 1, with x_{ij} being parents' level of education) using *lmer* in R with REML estimation. The estimated fixed component of the slope of parent education was positive and significant (2.90; $p < .05$), suggesting that students whose parents have higher levels of education tend to have higher math scores. However, this fixed effect estimate is conflated (*fixed conflation*) in that it implicitly reflects both a within-cluster fixed effect—i.e., the fixed component of a student's parents' education relative to their schoolmates—and a between-cluster fixed effect—i.e., the fixed component of the school's overall level of parent education. As for the random component of the slope, the slope variance was small and non-significant (1.29; LRT with a 50:50 mixture of $\chi^2_{df=1}$ and $\chi^2_{df=2}$ yielding $p > .05$). This might lead one to think that there is little across-cluster variability in the association of parent education and math scores. However, this estimate is also conflated (*random conflation*) in that it implicitly reflects both across-cluster slope heterogeneity in the association of parent education with math scores as well as intercept heteroscedasticity by school-mean parent education. As shown above via simulation, it is possible to get a small and non-significant conflated slope variance estimate even when there is actually substantial slope heterogeneity across clusters.

I next fit the (random-conflated) conventional random-slope contextual effect model by adding a fixed slope of school-mean parents' education (i.e., Equation 9; Table 1 Row 2). Though this model unconflates the fixed component, it still yields a conflated random component. In this the case, the conflated slope variance is roughly the same as it was in the previous model, and is still small and non-significant (1.51; LRT with a 50:50 mixture of $\chi^2_{df=1}$ and $\chi^2_{df=2}$ yielding $p > .05$). If a researcher were to fit this model in practice, they might feel that level-specific effects are effectively disaggregated based on the current literature in which only the fixed component is emphasized, but the slope variance is still implicitly reflecting both across-cluster slope heterogeneity and intercept heteroscedasticity.

Lastly, to fully unconflate, I fit the unconflated random slope contextual effect model (Equation (13); Table 1 row 3) and the heteroscedastic random-slope group-mean-centered model with a random between effect (Equation (17); Table 1 row 6). I present the results of both of these together in Table 3 to highlight their equivalencies (the table note explains how parameter estimates and standard errors from one model can be expressed in terms of those of the other model; corresponding analytic derivations for these equivalencies are in

Appendix C). By unconflating the fixed component, we see that both the within and between fixed effects are positive and significant, but that there is a contextual effect in that the between effect is larger. Specifically, the estimated within-effect is 2.85 and the estimated between-effect is 5.90, and hence the estimated contextual effect is 5.90 – 2.85 = 3.05. In other words, the overall school-average level of parent education is more predictive (in terms of slope magnitude) of math scores than is an individual student's parental education. By unconflating the random component, unlike in the random-conflated models, results indicate significant slope heterogeneity (2.31; LRT with a 50:50 mixture of $\chi^2_{df=2}$ and $\chi^2_{df=3}$ yielding $p < .05$ for both models in Table 3) and significant intercept heteroscedasticity (with the school-specific intercept variance for school j given as $15.38 + 23.90x_{\bullet j} + 10.07x^2_{\bullet j}$; and with a LRT with a 50:50 mixture of $\chi^2_{df=2}$ and $\chi^2_{df=3}$ yielding $p < .05$ for both models in Table 3). In other words, (a) certain schools had a stronger association between average parent education and math score than other schools, and (b) schools with either low or high average parent education had more variability in math scores.

Importantly, the results obtained by unconflating the random component are in sharp contrast to what was found in the widely used conventional random-slope models, wherein the slope variance was small and nonsignificant. Thus, similar to what was demonstrated via simulation, conflating the random component can obscure the two level-specific processes and can lead a researcher in practice to erroneously assume there is no across-cluster heterogeneity in the association of parent education with math scores. In this case, when fitting the widely used uncentered random slope MLM or the conventional random-slope contextual effect model, the conflated residuals are weighted heavily towards 0 in comparison to the u_{wj}'s.

Chapter IX

Discussion

In this dissertation I began by highlighting the current practice in which researchers commonly fit random-slope contextual effect models that are thought to disaggregate level-specific effects (Bliese & Britt, 2001; Deemer et al., 2017; Diez-Roux et al., 2000; Espelage et al., 2003; Fischer et al., 2004; Hoffman & Stawski, 2009; Kidwell et al., 1997; Lee, 2009; Lee & Bryk, 1989; Merlo et al., 2005; Poteat et al., 2007; Schempf & Kaufman, 2012; Titus, 2004). I then raised several issues with this common practice: (i) it is not well understood that MLMs can have conflation of *either* the fixed and/or random component of the slope of x_{ij}, and the possibilities for types of slope conflation had not previously been fully enumerated; (ii) there is little appreciation that the conventional random-slope contextual effect model (i.e., Equation (9)) or uncentered random-slope MLMs yield a conflated random component of the slope of x_{ij}; (iii) there are negative consequences of fitting random-conflated models that have not been explained or demonstrated, including making several kinds of erroneous interpretations and inferences; and (iv) researchers are in need of a full delineation of which MLM specifications yield unconflated random components, as well as recommendations for which to use for particular purposes. I addressed the first issue by explicating analytically the conflation of both the fixed and/or random component of the slope of x_{ij} and providing a taxonomy of random slope conflation that clarified how models can be fully conflated (i.e., both fixed and random components are conflated), partially conflated (i.e., only one of the components are conflated), or unconflated. I addressed the second issue by demonstrating analytically that the inclusion of a fixed slope of $x_{\bullet j}$ in the conventional random-slope contextual effect model effectively unconflates the fixed component but fails to unconflate the random component. I addressed the third issue by demonstrating via simulation the propensity for the widely used conventional random-slope contextual effect model to yield high rates of type I error, low power, and bias in random slope variance estimation as well as bias in fixed effect standard errors. Finally, I addressed the fourth issue by identifying a suite of unconflated random slope MLM specifications that appropriately disaggregate the random component, showing via simulation how all of the aforementioned problems are remedied by fitting these unconflated models, and providing recommendations regarding which unconflated model to use for particular research goals.

Future directions. Several limitations of this work serve as future directions. First, although the assumptions of the conventional random-slope contextual effect model (i.e., $\mathrm{var}(u_{wj}) = \mathrm{var}(u_{bj})$ and $\mathrm{corr}(u_{wj}, u_{bj}) = 1$), are highly restrictive and unlikely to hold in practice, they nonetheless could be tested. For instance, one can compare a conventional random-slope contextual effect model against an unconflated random-slope contextual effect model using a likelihood ratio test (with an adjusted reference distribution given that the null hypothesized value would lie on the boundary of the parameter space; Self & Liang, 1987; Stram & Lee, 1994, 1995) or with information criteria. Before formally recommending such a procedure, however, the utility of this model selection approach should be investigated via simulation. For instance, in future work the power to detect unequal variances and imperfect correlations of u_{wj} and u_{bj} could be investigated at finite sample sizes that are typically encountered in the social sciences and education. At the very least, I strongly recommend that researchers not assume without scrutiny that these restrictive assumptions hold in practice. A point made by Cronbach (1976) (in reference to fixed components but applicable to both fixed and random components) is relevant here: "No doubt group effects are negligible in some instruction and in some social processes, but even the investigator who prefers to deny their existence will be wise to allow his data to speak on the point. (p. 235)." In my opinion, the most straightforward approach to avoid random conflation is not to undertake (fallible) testing of these restrictive assumptions, but rather to simply avoid making the assumptions in the first place by instead fitting one of the fully unconflated models detailed here (Table 1 rows 3-6).

Second, although here I focused on random slope models with a single level-1 variable used to predict the outcome (e.g., by including as predictors the level-1 variable in uncentered form as well as in cluster-mean form), in models with random slopes for multiple predictors, each individual level-1 variable's slope can be characterized as fully conflated, partially conflated, or unconflated. Third, for simplicity here I posited linear relationships between each predictor and the outcome; however, a level-1 variable can have distinct nonlinear (e.g., quadratic) within-cluster and between-cluster relationships with the outcome. Future work can explicate the consequences of conflating such nonlinear level-specific relationships.

Fourth, although here I focused on the impact of random conflation on point estimation, significance testing, and standard error estimation, future work can explicate the impact on measures of effect size, such as R-squared (R^2). A recent framework of R^2 measures for multilevel models was developed by Rights and Sterba (2018), in which they provide novel measures but also analytically relate ten existing measures as special cases of their framework (e.g., Johnson, 2014; Kreft & de Leeuw, 1998; Nakagawa & Schielzeth, 2013; Snijders & Bosker, 1999; Vonesh & Chinchilli, 1997; Xu, 2003). Since conflation can distort either the fixed components or the random components of the slopes of level-1 predictors, the estimated proportion of variance attributable to level-1 predictors can similarly be distorted. Though a full explication of this impact is outside the scope of the current paper, I caution researchers to be wary of interpreting R^2 estimates from conflated models.

Fifth, for simplicity and to mirror the majority of methodological work on disaggregating level-specific effects, here I focused on using observed cluster means, $x_{\cdot j}$'s, as predictors for the MLM specifications in Table 1 rows 2-6. These models can be modified, however, by instead using latent cluster means via multilevel structural equation modeling (MSEM; e.g., Asparouhov & Muthén, 2006; Lüdtke et al., 2008; Preacher, Zyphur, & Zhang, 2010). This would similarly accomplish the disaggregation afforded by the use of observed cluster means, but could be further advantageous in avoiding bias when the observed cluster means are measured with meaningful amounts of error. Nonetheless, the use of MSEM over MLM comes with some tradeoffs; the former's reduction in bias is at the expense of greater variability in estimates, particularly at smaller sample sizes (e.g., Preacher et al., 2011). Future work can investigate the differential utility of either of these approaches with respect to disaggregating both level-specific fixed and level-specific random effects.

Sixth, in the simulation conditions investigated in this dissertation, all models had high rates of convergence, but in practice researchers may encounter convergence issues, particularly when random slope variances and/or sample sizes are small. Such issues could arise more often for the unconflated random-slope contextual effect model and the equivalent heteroscedastic random-slope group-mean-centered model, as these each involve estimating two random slopes, one of which is that of a level-2 predictor. Future research could explicitly investigate conditions that would be expected to yield higher rates of nonconvergence than those found in the current study. Should researchers encounter convergence issues with the heteroscedastic models, they could instead opt for a more parsimonious group-mean-centered model that assumes intercept homoscedasticity (Table 1 rows 4 or 5). Additionally, using Bayesian estimation might ease estimability (Browne & Draper, 2006) in this context, for instance, by using priors (e.g., half-Cauchy; Gelman, 2006) to shrink random effect variance estimates toward non-negative values (Depaoli & Clifton, 2015). For more general recommendations regarding remedies for convergence issues in MLM, see Brauer & Curtin (2017).

Finally, though here I focused on multilevel models that accommodate nested data structures via random effects, models that instead accommodate nested data by exclusively modeling the residual covariance structure among observations within-cluster (often termed marginal models) are becoming more popular in psychology applications (e.g., McNeish, Stapleton, & Silverman, 2017). If interest lies in modeling and assessing across-cluster random slope variability (the context assumed in this dissertation), such marginal models would be inappropriate. However, when research interest lies in only fixed effects, either marginal models or MLM could be used (for advantages and disadvantages of these approaches, see, e.g., Gardiner, Luo, & Roman, 2009; McNeish et al. 2017). Future research could investigate the utility of marginal models in terms of fixed effects' standard error accuracy across the generating conditions investigated in this dissertation. It is expected, however, that marginal models would be less efficient than MLM when the generating model contains random slopes, as the former could not as readily and parsimoniously represent such data generating structures.

Conclusion. Despite the predominant focus in the methodological literature on the conflation of level-specific fixed effects, I have shown here that the conflation of random effects can similarly lead to meaningful distortion in multilevel models. It is my hope that this work will help researchers understand the distinction between random effects associated with level-1 variables vs. those associated level-2 variables, and will make clear the adverse impact of conflating the two. I further hope that this work will ultimately help encourage the use of unconflated models in practice.

References

Algina, J., & Swaminathan, H. (2011). Centering in two-level nested designs. In J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 285–312). New York: Taylor and Francis.

Asparouhov, T., & Muthén, B. (2006). *Constructing covariates in multilevel regression* (Mplus Web Note No. 11). Retrieved from http://www.statmodel.com.

Baldwin, S. A., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology*, *75*, 842.

Bliese, P. D., & Britt, T. W. (2001). Social support, group consensus and stressor–strain relationships: Social context matters. *Journal of Organizational Behavior*, *22*, 425-436.

Boateng, F. D. (2016). Neighborhood-level effects on trust in the police: A multilevel analysis. *International Criminal Justice Review, 26*, 217-236.

Brauer, M., & Curtin, J. J. (2017). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods, 23*, 389-411.

Browne, W. J., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, *1*, 473-514.

Cronbach, L. J. (1976). *Research on classrooms and schools: Formulation of questions, design, and analysis.* Unpublished manuscript, Stanford University, Stanford Evaluation Consortium, School of Education.

Curran, P. J., Lee, T., Howard, A. L., Lane, S., & MacCallum, R. A. (2012). Disaggregating within-person and between-person effects in multilevel and structural equation growth models. In J. R. Harring & G. R. Hancock (Eds.), *Advances in longitudinal methods in the social and behavioral sciences* (pp. 217–253). Charlotte, NC: Information Age.

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*, 583-619.

Deemer, E. D., Marks, L. R., & Miller, K. A. (2017). Peer science self-efficacy: A proximal contextual support for college students' science career intentions. *Journal of Career Assessment*, *25*, 537-551.

Demidenko, E. (2004). *Mixed models: Theory and applications*. Hoboken, NJ: Wiley.

Depaoli, S., & Clifton, J. P. (2015). A Bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*, 327-351.

Diez-Roux, A. V., Link, B. G., & Northridge, M. E. (2000). A multilevel analysis of income inequality and cardiovascular disease risk factors. *Social Science & Medicine, 50*, 673-687.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*, 121-138.

Espelage, D. L., Holt, M. K., & Henkel, R. R. (2003). Examination of peer–group contextual effects on aggression during early adolescence. *Child Development*, *74*, 205-220.

Fischer, A. H., Rodriguez Mosquera, P. M., van Vianen, A. E. M., & Manstead, A. S. R. (2004). Gender and culture differences in emotion. *Emotion*, 4, 87–94.

Gardiner, J. C., Luo, Z., & Roman, L. A. (2009). Fixed effects, random effects and GEE: what are the differences?. *Statistics in Medicine, 28*, 221-239.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis, 1*, 515-534.

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). New York, NY: Wiley.

Hatcher, R. L., & Barends, A. W. (2006). How a return to theory could help alliance research. *Psychotherapy: Theory, Research, Practice, Training, 43*, 292–299.

Hedeker, D., & Gibbons, R. (2006). *Longitudinal data analysis*. New York, NY: Wiley.

Henry, K. L., & Slater, M. D. (2007). The contextual effect of school attachment on young adolescents' alcohol use. *Journal of School Health, 77*, 67-74.

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 24*, 623-641.

Hoffman, L. (2015). Longitudinal analysis: Modeling within-person fluctuation and change. New York, NY: Routledge.

Hoffman, L., & Stawski, R. S. (2009). Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. *Research in Human Development, 6*, 97-120.

Hox, J. J. (2010). *Multilevel analysis. Techniques and applications*. (2nd Ed). New York: Routledge.

Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods in Ecology and Evolution, 5*, 944-946.

Kidwell Jr, R. E., Mossholder, K. W., & Bennett, N. (1997). Cohesiveness and organizational citizenship behavior: A multilevel analysis using work groups and individuals. *Journal of Management*, *23*, 775-793.

Kreft, I. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.

Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effects of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research, 30*, 1-21.

Lee, C. C. (2009). Hierarchical linear modeling to explore the influence of satisfaction with public facilities on housing prices. *International Real Estate Review*, *12*, 252-272.

Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, *62*, 172-192.

Longford, N. (1989). To center or not to center. *Multilevel Modelling Newsletter*, *1*, 7.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: a new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*, 203.

Mallinckrodt, B. (2000). Attachment, social competencies, social support, and interpersonal process in psychotherapy. *Psychotherapy Research, 10*, 239 –266.

McNeish, D., Stapleton, L.M., & Silverman, R.D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods, 22*, 114–140.

Merlo, J., Chaix, B., Yang, M., Lynch, J., & Råstam, L. (2005). A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *Journal of Epidemiology & Community Health*, *59*, 1022-1029.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133-

142.

Paccagnella, O. (2006). Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. *Evaluation Review, 30*, 66-85.

Plewis, I. (1989). Comment on "centering" predictors. *Multilevel Modelling Newsletter*, *1*, 6 & 11.

Poteat, V. P., Espelage, D. L., & Green Jr, H. D. (2007). The socialization of dominance: Peer group contextual effects on homophobic and dominance attitudes. *Journal of Personality and Social Psychology*, *92*, 1040-1050.

Preacher, K. J. (2011). Multilevel SEM strategies for evaluating mediation in three-level data. *Multivariate Behavioral Research, 46*, 691-731.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling*, *18,* 161-182.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods, 15,* 209.

Raudenbush, S. W. (1989). "Centering" predictors in multilevel analysis: Choices and consequences. *Multilevel Modelling Newsletter, 1,* 10-12.

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

Rights, J. D., & Sterba, S. K. (2018). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods.* doi:10.1037/met0000184.

Rights, J. D., & Sterba, S. K. (2016). The relationship between multilevel models and nonparametric multilevel mixture models: Discrete approximation of intraclass correlation, random coefficient distributions, and residual heteroscedasticity. *British Journal of Mathematical and Statistical Psychology*, *69*, 316-343.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*, 351-357.

Scheipl, F., and Bolker, B. (2013), *RLRsim: Exact (restricted) likelihood ratio tests for mixed and additive models.* R package version 2.0-12.

Schempf, A. H., & Kaufman, J. S. (2012). Accounting for context in studies of health inequalities: a review and comparison of analytic approaches. *Annals of Epidemiology*, *22*, 683-690.

Self, S.G., & Liang, H.Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, *82*, 606–610.

Snijders, T. A. B., & Berkhof, J. (2008). Diagnostic checks for multilevel models. In J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 141-176). New York, NY: Springer.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London: Sage.

Stram, D. O., & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics, 50*, 1171–1177.

Stram, D. O., & Lee, J. W. (1995). Corrections: Variance components testing in the longitudinal mixed effects model. *Biometrics, 51*, 1196.

Titus, M. A. (2004). An examination of the influence of institutional context on student persistence at 4-year colleges and universities: A multilevel approach. *Research in Higher Education*, *45*, 673-699.

Verbeke, G., & Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics & Data Analysis, 23*, 541-556.

Vonesh, E. F., & Chinchilli, V. M. (1997). *Linear and nonlinear models for the analysis of repeated measurements.* New York: Marcel Dekker.

Wang, L. P., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, *20,* 63-83.

Wu, Y. W. B., & Wooldridge, P. J. (2005). The impact of centering first-level predictors on individual and contextual effects in multilevel data analysis. *Nursing Research*, *54,* 212-216.

Xu, R. H. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine, 22,* 3527-3541.

Table 1. Suite of random slope models characterized by full conflation, partial conflation, or no conflation

Manuscript Equation #	**Fixed slope components for:**	**Random slope components for:**	**Conflation?**	**Commonly used?**	**Previously Recommended?**	**Currently Recommended Here?**	**Name of model specification**
			Uncentered (or grand-mean-centered) model				
Eqn (7)	x_{ij}	x_{ij}	fixed+random conflation	Yes		No	uncentered random-slope MLM
			Contextual effect models				
Eqn (9)	x_{ij} and $x_{\bullet j}$	x_{ij}	random conflation	Yes	*Hox (2010); Kreft, de Leeuw, & Aiken (1995); Longford (1989); Plewis (1989); Snijders & Bosker (2012)*	No	conventional random-slope contextual effect MLM
Eqn (13)	x_{ij} and $x_{\bullet j}$	x_{ij} and $x_{\bullet j}$	unconflated	No		Yes	unconflated random-slope contextual effect MLM
			Group-mean-centered models				
Eqn (15)	$(x_{ij} - x_{\bullet j})$	$(x_{ij} - x_{\bullet j})$	unconflated	Yes		Yes	random-slope group-mean-centered MLM
Eqn (16)	$(x_{ij} - x_{\bullet j})$ and $x_{\bullet j}$	$(x_{ij} - x_{\bullet j})$	unconflated	Yes	*Raudenbush (1989); Raudenbush & Bryk (2002); Wang & Maxwell (2015); Hoffman & Gavin (1998)*	Yes*	random-slope group-mean-centered MLM w/ fixed between effect
Eqn (17)	$(x_{ij} - x_{\bullet j})$ and $x_{\bullet j}$	$(x_{ij} - x_{\bullet j})$ and $x_{\bullet j}$	unconflated	No		Yes	heteroscedastic random-slope group-mean-centered MLM w/ btwn. eff.

Notes: MLM=multilevel model. *=Note that this model assumes intercept homoscedasticity; the presence of intercept heteroscedasticity can bias inference about level-2 fixed effects (see text for details).

Table 2. Testing the random slope variance of a level-1 predictor using the conventional random-slope contextual effect model vs. the unconflated random-slope models: Power and type I error

	Power	**Type I error rate**
Conventional random-slope contextual effect model	.69	.13
Unconflated random-slope contextual effect model	.87	.03
Random-slope group-mean-centered model w/o $x_{\cdot j}$	.92	.04
Random-slope group-mean-centered model w/ fixed between effect	.91	.04
Heteroscedastic random-slope group-mean-centered model w/ between effect	.87	.03

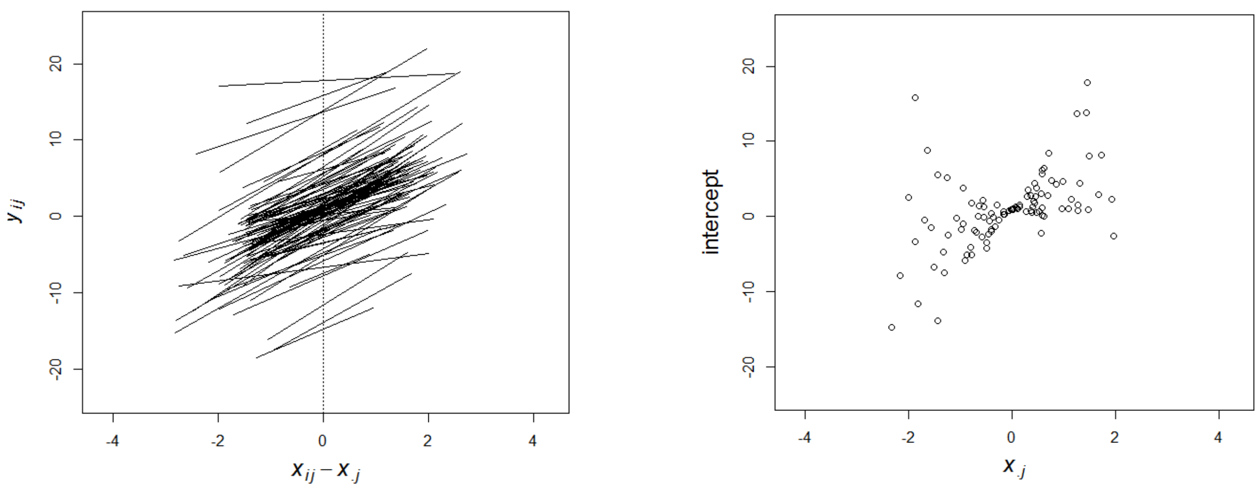
Table 3. Empirical example results for unconflated random-slope models

Unconflated random-slope contextual effect model **(Eq. 13)**		***Heteroscedastic random-slope group-mean-centered model*** **(Eq. 17)**	
Fixed effect estimates (SE)			
intercept ($\hat{\gamma}_{00}$)	52.38 (1.08)	intercept ($\hat{\gamma}_{00}$)	52.38 (1.08)
slope of parent education ($\hat{\gamma}_{10}$)	2.85 (0.49)	slope of school-mean-centered parent education ($\hat{\gamma}_{10}$)	2.85 (0.49)
slope of school-mean parent education ($\hat{\gamma}_{01}$)	3.05 (1.27)	slope of school-mean parent education ($\hat{\gamma}_{01}$)	5.90 (1.24)
Variance component estimates†			
$\hat{\tau}_{00}$	15.38	$\hat{\tau}_{00}$	15.38
parent education slope resid. variance ($\hat{\tau}_{11}$)	2.31	school-mean-centered parent ed. slope resid. variance ($\hat{\tau}_{11}$)	2.31
school-mean parent education slope resid. variance ($\hat{\tau}_{22}$)	8.46	school-mean parent education slope resid. variance ($\hat{\tau}_{22}$)	10.07
covariance of u_{0j} & parent education slope resid. ($\hat{\tau}_{10}$)	0.71	covariance of u_{0j} & school-mean-centered parent education slope resid. ($\hat{\tau}_{10}$)	0.71
covariance of u_{0j} & school-mean parent ed. slope resid. ($\hat{\tau}_{20}$)	11.18	covariance of u_{0j} & school-mean parent ed. slp. resid. ($\hat{\tau}_{20}$)	11.88
covariance of parent education slope resid. with school-mean parent education slope resid. ($\hat{\tau}_{21}$)	-0.35	covariance of school-mean-centered parent ed. slope resid. & school-mean parent ed. slope resid. ($\hat{\tau}_{21}$)	1.96
level-1 residual variance ($\hat{\sigma}^2$)	70.95	level-1 residual variance ($\hat{\sigma}^2$)	70.95

Note. $\hat{\gamma}_{01}$, $\hat{\tau}_{20}$, $\hat{\tau}_{21}$, and $\hat{\tau}_{22}$ from the heteroscedastic random-slope group-mean-centered model are equal to $(\hat{\gamma}_{10}+\hat{\gamma}_{01})$, $(\hat{\tau}_{10}+\hat{\tau}_{20})$, $(\hat{\tau}_{11}+\hat{\tau}_{21})$, and $(\hat{\tau}_{11}+\hat{\tau}_{22}+2\hat{\tau}_{21})$, respectively, from the unconflated random-slope contextual effect model. Additionally, the standard error of $\hat{\gamma}_{01}$ in the heteroscedastic random-slope group-mean-centered model is equal to $se_{\hat{\gamma}^*_{01}} = \sqrt{se^2_{\hat{\gamma}_{10}} + se^2_{\hat{\gamma}_{01}} + 2\operatorname{cov}(\hat{\gamma}_{10},\hat{\gamma}_{01})} = 1.24$ (with $\operatorname{cov}(\hat{\gamma}_{10},\hat{\gamma}_{01}) = -0.16$) from the unconflated random-slope contextual effect model (see Appendix C). All fixed effects were significant (*t*-test with alpha = .05). †=All random effect variances were significant using mixture likelihood ratio tests described in the Empirical Example section. Resid.=Residual. Eq.=Equation.

Figure 1. Interpretation of a random component of a level-1 predictor vs. that of a level-2 predictor

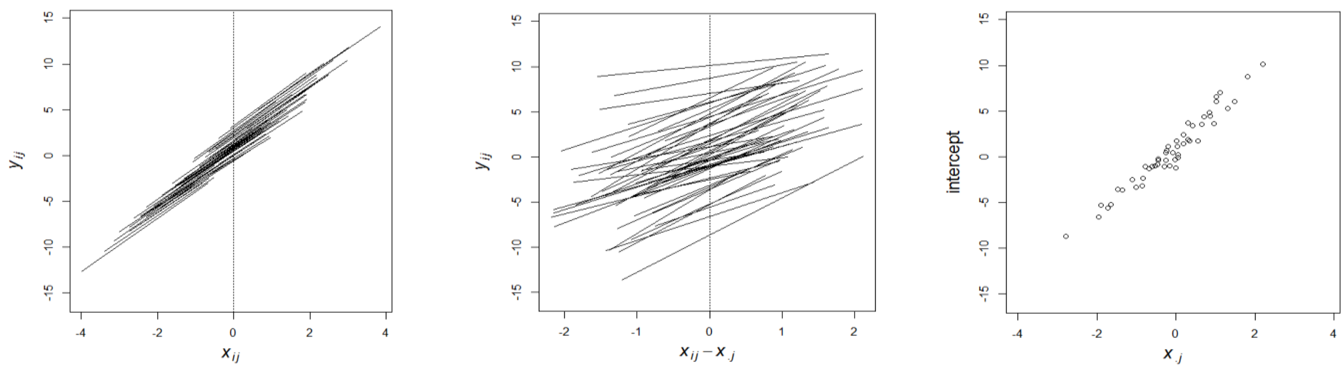
Panel A: A random component of a purely level-1 predictor reflects across-cluster *slope heterogeneity*	**Panel B**: A random component of a purely level-2 predictor reflects across-cluster *intercept heteroscedasticity*



Notes. In Panel A, slope heterogeneity is reflected in the fact that the within-cluster effect of $x_{ij} - x_{\bullet j}$ depends on cluster membership. In Panel B, heteroscedastic intercept variance is reflected in the fact that cluster-specific intercepts are more variable at the extremes of $x_{\bullet j}$.

Figure 2. Illustrating Erroneous Inference #1 from the conventional random-slope contextual effect model: Concluding there is no random slope variation when it exists

Panel A: Conventional random-slope contextual effect model results	**Panel B**: Actual within-cluster relations (random slope variation)	**Panel C**: Actual between relations (intercept homoscedasticity)



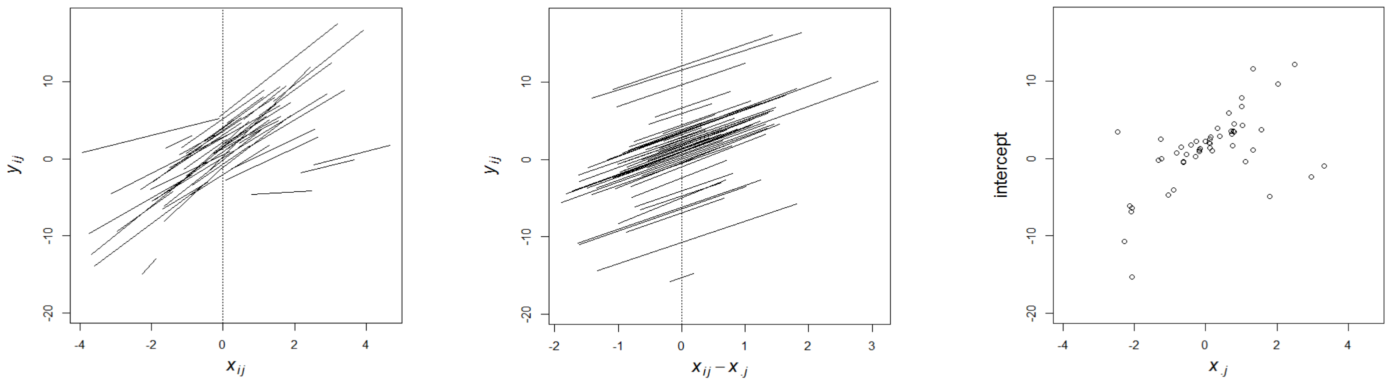
$\hat{\tau}_{11} = \text{var}(u_{1j}) = 0.04; ns$

Researcher would erroneously conclude: no random slope variation

Note. The significance test for τ_{11} uses a χ^2 mixture likelihood ratio test [LRT] with the null reference distribution being a 50:50 mixture of $\chi^2_{df=1}$ and $\chi^2_{df=2}$ (see Footnote 10).

Figure 3. Illustrating Erroneous Inference #2 from the conventional random-slope contextual effect model: Concluding there is random slope variation when it does not exist

Panel A: Conventional random-slope contextual effect model results	**Panel B**: Actual within-cluster relations (fixed slopes)	**Panel C**: Actual between-cluster relations (intercept heteroscedasticity)

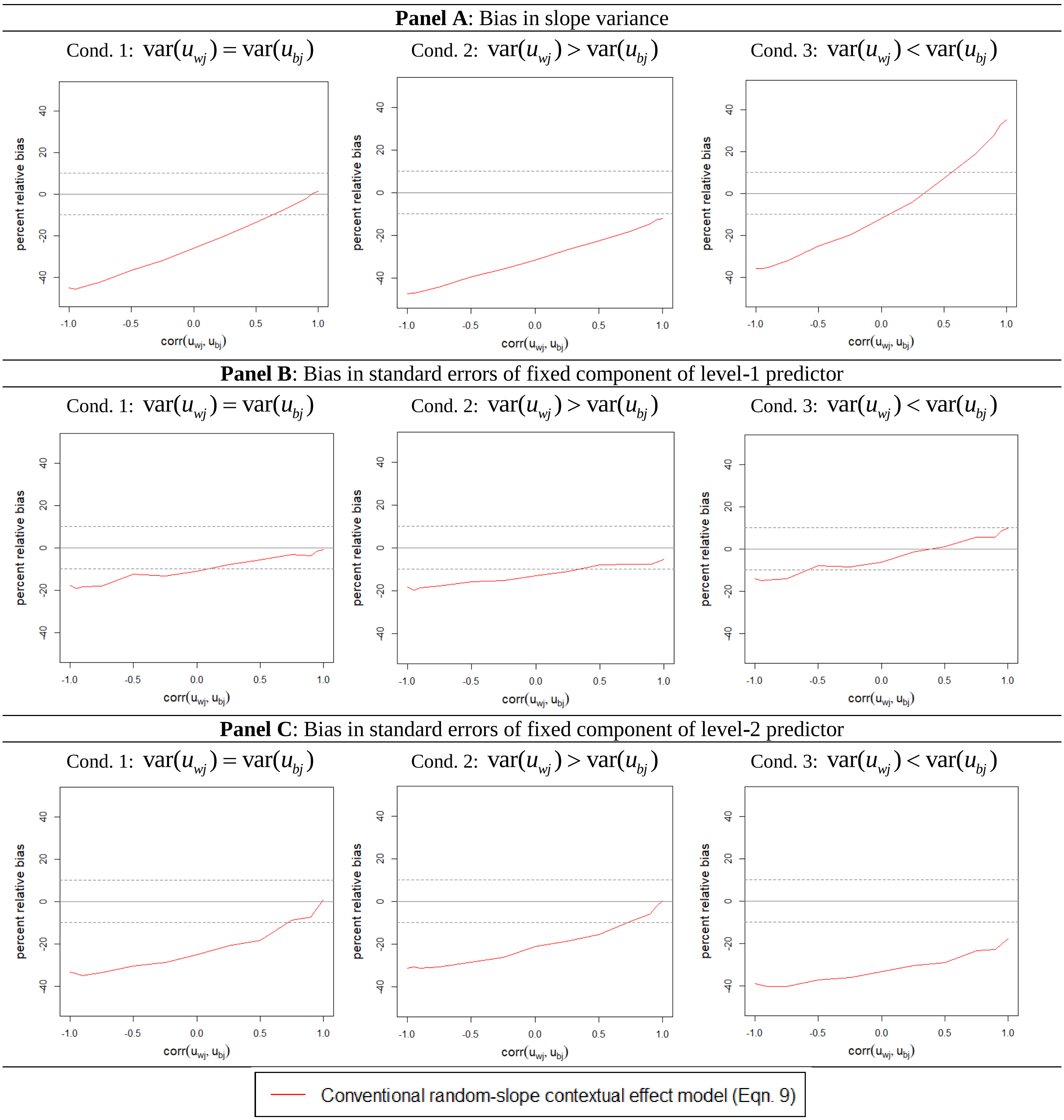


$\hat{\tau}_{11} = \text{var}(u_{1j}) = 3.01*$

Researcher would erroneously conclude: random slope variation

Note. The significance test for τ_{11} uses a χ^2 mixture likelihood ratio test [LRT] with the null reference distribution being a 50:50 mixture of $\chi^2_{df=1}$ and $\chi^2_{df=2}$ (see Footnote 10).

Figure 4. Bias in random slope variance estimates and fixed effect standard errors for the conventional random-slope contextual effect model

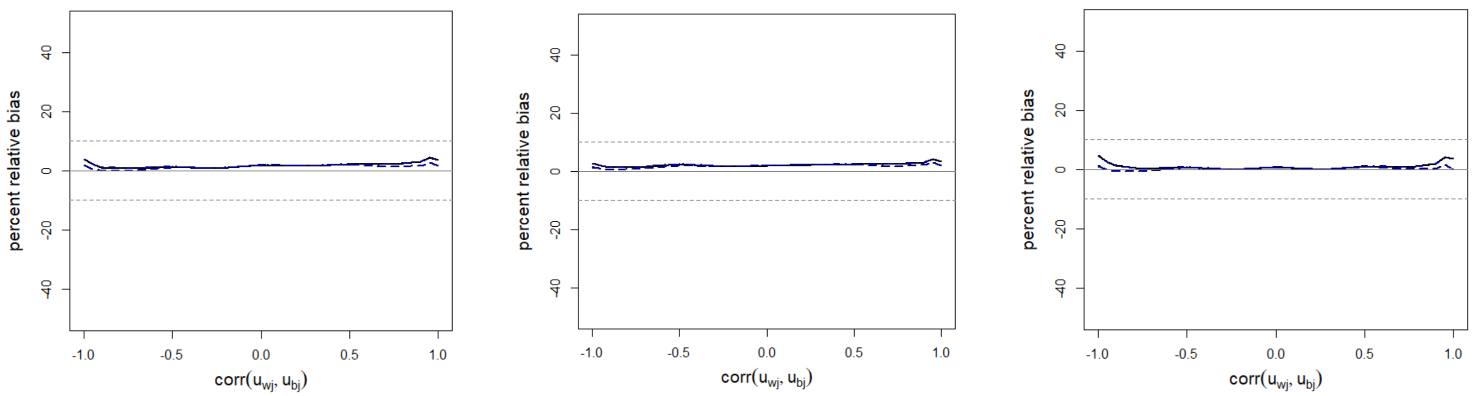


Notes. Cond.=Condition. The three horizontal (grey) reference lines denote where absolute relative bias is 10%, 0%, and -10%; values outside of this range conventionally indicate meaningfully substantial bias.

Figure 5. Bias in random slope variance estimates and fixed effect standard errors for the unconflated random-slope models

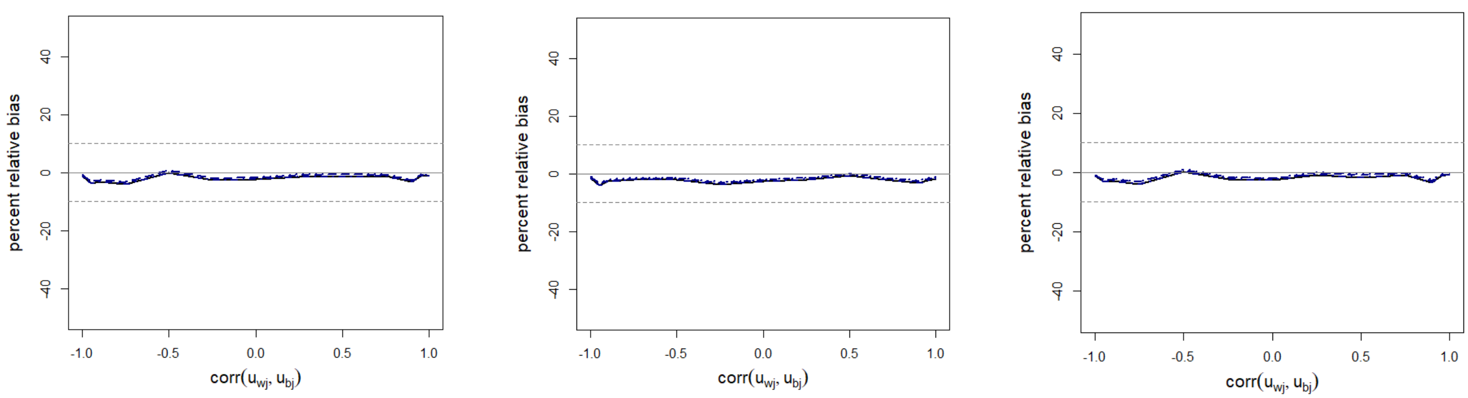
Panel A: Bias in slope variance

Cond. 1: $\mathrm{var}(u_{wj}) = \mathrm{var}(u_{bj})$ Cond. 2: $\mathrm{var}(u_{wj}) > \mathrm{var}(u_{bj})$ Cond. 3: $\mathrm{var}(u_{wj}) < \mathrm{var}(u_{bj})$



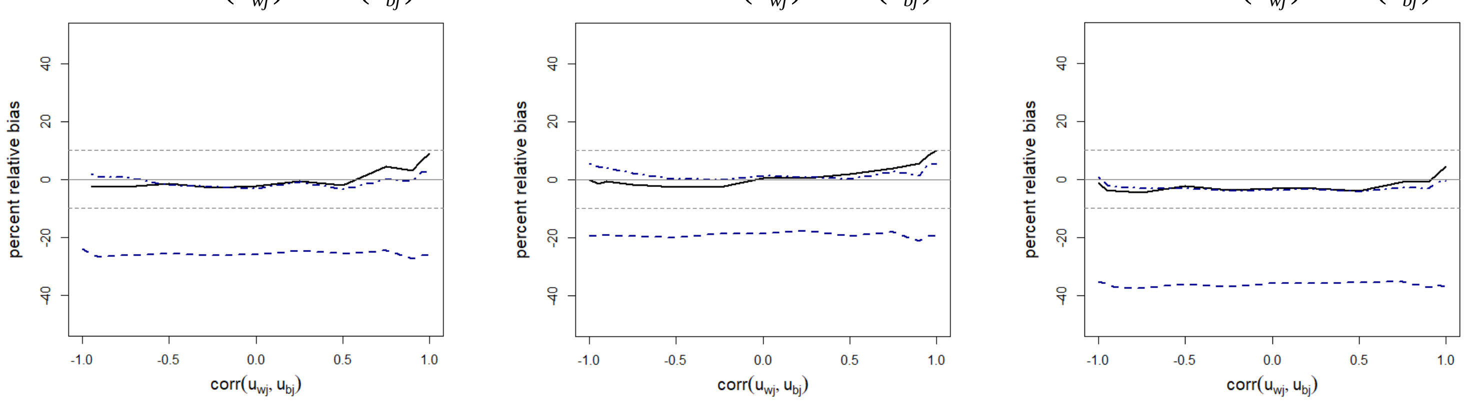
Panel B: Bias in standard errors of fixed component of level-1 predictor

Cond. 1: $\mathrm{var}(u_{wj}) = \mathrm{var}(u_{bj})$ Cond. 2: $\mathrm{var}(u_{wj}) > \mathrm{var}(u_{bj})$ Cond. 3: $\mathrm{var}(u_{wj}) < \mathrm{var}(u_{bj})$



Panel C: Bias in standard errors of fixed component of level-2 predictor

Cond. 1: $\mathrm{var}(u_{wj}) = \mathrm{var}(u_{bj})$ Cond. 2: $\mathrm{var}(u_{wj}) > \mathrm{var}(u_{bj})$ Cond. 3: $\mathrm{var}(u_{wj}) < \mathrm{var}(u_{bj})$



Unconflated random-slope contextual effect model (Eqn. 13)
Random-slope group-mean-centered model (Eqn. 15)
Random-slope group-mean-centered model w/ fixed between effect (Eqn. 16)
Heteroscedastic random-slope group-mean-centered model w/ between effect (Eqn. 17)

Notes. Cond.=Condition. The three horizontal (grey) reference lines denote where absolute relative bias is 10%, 0%, and -10%; values outside of this range conventionally indicate meaningfully substantial bias.

Appendix A: Derivation of the heteroscedastic intercept variance in Equation 11 that is implied by the conventional random-slope contextual effect model

Here I show that the heteroscedastic intercept variance for the conventional random-slope contextual effect model is given by the quadratic function of $x_{\bullet j}$ in Equation 11. First I show how the conventional random-slope contextual effect model equation given in Equation 10 can be expressed in terms of cluster-specific intercepts and slopes:

$$y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij} - x_{\bullet j}) + e_{ij} \tag{A1}$$

Here β_{0j} is the cluster-specific intercept, and β_{1j} the cluster-specific slope of $(x_{ij} - x_{\bullet j})$. Defining these as

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + (\gamma_{10} + \gamma_{01})x_{\bullet j} + u_{0j} + u_{1j}x_{\bullet j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \tag{A2}$$

yields Equation 10.

Next, noting that, in general, a model's random intercept variance can be expressed as the variance of β_{0j} conditional on predictors (Goldstein, 2011; Rights & Sterba, 2016; Snijders & Bosker, 2012), I compute the random intercept variance of the conventional random-slope contextual effect model as:

$$\begin{aligned} \mathrm{var}(\beta_{0j} \mid x_{\bullet j}) &= \mathrm{var}(\gamma_{00} + (\gamma_{10} + \gamma_{01})x_{\bullet j} + u_{0j} + u_{1j}x_{\bullet j} \mid x_{\bullet j}) \\ &= \mathrm{var}(u_{0j} + u_{1j}x_{\bullet j} \mid x_{\bullet j}) \\ &= \mathrm{var}(u_{0j}) + 2\,\mathrm{cov}(u_{1j}x_{\bullet j}, u_{0j} \mid x_{\bullet j}) + \mathrm{var}(u_{1j}x_{\bullet j} \mid x_{\bullet j}) \\ &= \mathrm{var}(u_{0j}) + 2\,\mathrm{cov}(u_{1j}, u_{0j})x_{\bullet j} + \mathrm{var}(u_{1j})x_{\bullet j}^2 \\ &= \tau_{00} + 2\tau_{10}x_{\bullet j} + \tau_{11}x_{\bullet j}^2 \end{aligned} \tag{A3}$$

This is equal to τ_{22j} given in Equation 11.

Appendix B: The conventional random-slope contextual effect model (Equation 9) is nested within the unconflated random-slope contextual effect model (Equation 13)

Here I show that the conventional random-slope contextual effect model (manuscript Equation 9) is nested within the proposed unconflated random-slope contextual effect model with distinct fixed and random components for $x_{ij} - x_{\bullet j}$ and $x_{\bullet j}$ (manuscript Equation 13) via two constraints on Equation 12: $\mathrm{var}(u_{wj}) = \mathrm{var}(u_{bj})$ and $\mathrm{corr}(u_{wj}, u_{bj}) = 1$. I will do this by showing that the unconflated random-slope contextual effect model is equivalent to the conventional random-slope contextual effect model when imposing these constraints.

I first show that $\mathrm{var}(u_{wj}) = \mathrm{var}(u_{bj})$ and $\mathrm{corr}(u_{wj}, u_{bj}) = 1$ together implies $u_{wj} = u_{bj}$. Note first that, by definition:

$$\begin{aligned} E[u_{wj} - u_{bj}] &= E[u_{wj}] - E[u_{bj}] \\ &= 0 - 0 \\ &= 0 \end{aligned} \tag{B1}$$

Next, if the two constraints hold:

$$\begin{aligned} \mathrm{var}(u_{bj} - u_{wj}) &= \mathrm{var}(u_{bj}) + \mathrm{var}(u_{wj}) - 2\,\mathrm{cov}(u_{bj}, u_{wj}) \\ &= \mathrm{var}(u_{bj}) + \mathrm{var}(u_{wj}) - 2\sqrt{\mathrm{var}(u_{bj})\,\mathrm{var}(u_{wj})}\,\mathrm{corr}(u_{bj}, u_{wj}) \\ &= \mathrm{var}(u_{bj}) + \mathrm{var}(u_{wj}) - 2\sqrt{\mathrm{var}(u_{bj})\,\mathrm{var}(u_{wj})} \\ &= \left(\sqrt{\mathrm{var}(u_{bj})} - \sqrt{\mathrm{var}(u_{wj})}\right)^2 \\ &= \left(\sqrt{\mathrm{var}(u_{bj})} - \sqrt{\mathrm{var}(u_{bj})}\right)^2 \\ &= 0 \end{aligned} \tag{B2}$$

Together, Equations B1 and B2 imply that $u_{wj} - u_{bj} = 0$ for all *j*, which implies that $u_{wj} = u_{bj}$ for all *j*. We can thus set the u_{wj} and u_{bj} terms in the unconflated model in Equation 12 generically to u_{1j}, which yields the following model:

$$y_{ij} = \gamma_{00} + \gamma_w (x_{ij} - x_{\bullet j}) + \gamma_b x_{\bullet j} + u_{1j}(x_{ij} - x_{\bullet j}) + u_{1j} x_{\bullet j} + u_{0j} + e_{ij}$$

$$e_{ij} \sim N(0, \sigma^2)$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{10} \\ \tau_{10} & \tau_{11} \end{bmatrix} \right) \tag{B3}$$

This is equivalent to the (random-conflated) conventional random-slope contextual effect model, as shown in parameterization given in manuscript Equation 10. With manuscript Equation 10, we can simply set $\gamma_{10} = \gamma_w$ and $\gamma_{01} = \gamma_b - \gamma_w$, and the expression is the same as Equation B3.

Appendix C: The unconflated random-slope contextual effect model (Equation 13) is equivalent to the heteroscedastic random-slope group-mean-centered model (Equation 17)

Here I show that the unconflated random-slope contextual effect model (manuscript Equation 13) is equivalent to the heteroscedastic group-mean-centered model (manuscript Equation 17), and the two are different parameterizations of each other. The unconflated random-slope contextual effect model can be re-expressed as

$$\begin{aligned} y_{ij} &= \gamma_{00} + \gamma_{10}x_{ij} + u_{1j}x_{ij} + \gamma_{01}x_{\bullet j} + u_{2j}x_{\bullet j} + u_{0j} + e_{ij} \\ &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\bullet j} + x_{\bullet j}) + u_{1j}(x_{ij} - x_{\bullet j} + x_{\bullet j}) + \gamma_{01}x_{\bullet j} + u_{2j}x_{\bullet j} + u_{0j} + e_{ij} \\ &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\bullet j}) + \gamma_{10}x_{\bullet j} + u_{1j}(x_{ij} - x_{\bullet j}) + u_{1j}x_{\bullet j} + \gamma_{01}x_{\bullet j} + u_{2j}x_{\bullet j} + u_{0j} + e_{ij} \\ &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\bullet j}) + (\gamma_{10} + \gamma_{01})x_{\bullet j} + u_{1j}(x_{ij} - x_{\bullet j}) + (u_{1j} + u_{2j})x_{\bullet j} + u_{0j} + e_{ij} \end{aligned} \tag{C1}$$

Letting γ_{01}^{*} and u_{2j}^{*} denote, respectively, the fixed and random component of the slope of $x_{\bullet j}$ from the heteroscedastic group-mean-centered model, I define these parameters in terms of unconflated random-slope contextual effect model parameters as $\gamma_{01}^{*} = \gamma_{10} + \gamma_{01}$ and $u_{2j}^{*} = u_{1j} + u_{2j}$ to show the equivalence of the reduced-form expressions in the first lines of Equations 13 and 17 (all other terms between the models on these lines are the exact same). This implies that the standard error of $\hat{\gamma}_{01}^{*}$ can be expressed as the square root of a function of the asymptotic (i.e. across repeated sampling) variances and covariances of the estimates from the unconflated random-slope contextual effect model, as such:

$$\begin{aligned} se_{\hat{\gamma}_{01}^{*}} &= \sqrt{\mathrm{var}(\hat{\gamma}_{01}^{*})} \\ &= \sqrt{\mathrm{var}(\hat{\gamma}_{10} + \hat{\gamma}_{01})} \\ &= \sqrt{\mathrm{var}(\hat{\gamma}_{10}) + \mathrm{var}(\hat{\gamma}_{01}) + 2\,\mathrm{cov}(\hat{\gamma}_{10}, \hat{\gamma}_{01})} \\ &= \sqrt{se_{\hat{\gamma}_{10}}^{2} + se_{\hat{\gamma}_{01}}^{2} + 2\,\mathrm{cov}(\hat{\gamma}_{10}, \hat{\gamma}_{01})} \end{aligned} \tag{C2}$$

Further, I show the equivalence of the random effect covariances by first letting τ^*_{20}, τ^*_{21}, and τ^*_{22} denote variance and covariance components from the heteroscedastic random-slope group-mean-centered model, and then expressing the random effect covariance structure of this model as

$$\begin{bmatrix} u_{0j} \\ u_{1j} \\ u^*_{2j} \end{bmatrix} = \begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{1j}+u_{2j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ \tau^*_{20} & \tau^*_{21} & \tau^*_{22} \end{bmatrix} \right) \tag{C3}$$

with u_{0j}, u_{1j}, τ_{00}, τ_{10}, and τ_{11} being the same for either model. I next express this random effect covariance structure purely in terms of the unconflated random-slope contextual effect model parameters by considering that

$$\begin{aligned} \tau^*_{20} &= \mathrm{cov}(u_{1j}+u_{2j}, u_{0j}) \\ &= \mathrm{cov}(u_{1j}, u_{0j}) + \mathrm{cov}(u_{2j}, u_{0j}) \\ &= \tau_{10} + \tau_{20} \end{aligned} \tag{C4}$$

and

$$\begin{aligned} \tau^*_{21} &= \mathrm{cov}(u_{1j}+u_{2j}, u_{1j}) \\ &= \mathrm{cov}(u_{1j}, u_{1j}) + \mathrm{cov}(u_{2j}, u_{1j}) \\ &= \mathrm{var}(u_{1j}) + \mathrm{cov}(u_{2j}, u_{1j}) \\ &= \tau_{11} + \tau_{21} \end{aligned} \tag{C5}$$

and

$$\begin{aligned} \tau^*_{22} &= \mathrm{var}(u_{1j}+u_{2j}) \\ &= \mathrm{var}(u_{1j}) + \mathrm{var}(u_{2j}) + 2\,\mathrm{cov}(u_{1j}, u_{2j}) \\ &= \tau_{11} + \tau_{22} + 2\tau_{21} \end{aligned} \tag{C6}$$

Thus,

$$\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{1j}+u_{2j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ \tau_{10}+\tau_{20} & \tau_{11}+\tau_{21} & \tau_{11}+\tau_{22}+2\tau_{21} \end{bmatrix}\right) \tag{C7}$$

Hence, the heteroscedastic random-slope group-mean-centered model parameters can be equivalently expressed in terms of unconflated random-slope contextual effect model parameters, and the two models are different parameterizations of each other.