Understanding the consequences of collinearity for multilevel models: The importance of disaggregation across levels

Haley E. Yaremych & Kristopher J. Preacher

Department of Psychology & Human Development, Vanderbilt University

Correspondence concerning this article should be addressed to:
Haley E. Yaremych
Department of Psychology & Human Development, Vanderbilt University
PMB 552, 230 Appleton Place
Nashville, TN 37203-5721
haley.e.yaremych@vanderbilt.edu

Disclosure statement: The authors report there are no competing interests to declare.

Funding: This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1937963. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Data availability statement: Code for data generation is available in the Online Code Supplement.

Abstract

In multilevel models, disaggregating predictors into level-specific parts (typically accomplished via centering) benefits parameter estimates and their interpretations. However, the importance of level-specificity has been sparsely addressed in multilevel literature concerning collinearity. In this study, we develop novel insights into the interactivity of centering and collinearity in multilevel models. After integrating the broad literatures on centering and collinearity, we review level-specific and conflated correlations in multilevel data. Next, by deriving formal relationships between predictor collinearity and multilevel model estimates, we demonstrate how the consequences of collinearity change across different centering specifications and identify data characteristics that may exacerbate or mitigate those consequences. We show that when all or some level-1 predictors are uncentered, slope estimates can be greatly biased by collinearity. Disaggregation of all predictors eliminates the possibility that fixed effect estimates will be biased due to collinearity alone; however, under some data conditions, collinearity is associated with biased standard errors and random effect (co)variance estimates. Finally, we illustrate the importance of disaggregation for diagnosing collinearity in multilevel data and provide recommendations for the use of level-specific collinearity diagnostics. Overall, the necessity of disaggregation for identifying and managing collinearity's consequences in multilevel models is clarified in novel ways.

Keywords: *collinearity; multicollinearity; correlation; multilevel models; hierarchical linear models; centering*

Collinearity

Collinearity is broadly defined as interdependency, redundancy, or shared variance among predictors in a data set. *Exact collinearity* exists if there are one or more perfect linear relationships among the predictors, however this is rarely observed in applied settings (Silvey, 1969; Willan & Watts, 1978). Therefore, in practice, collinearity refers to *near collinearity,* which exists when there are nearly or approximately perfect linear relationships among predictors (Belsley, 1991; Gunst & Mason, 1977). Instead of declaring that collinearity is present or not present in a data set, we should acknowledge that it always exists to some degree, and our goal is to understand the severity of its effects (Farrar & Glauber, 1967; Yu et al., 2015). These ideas and definitions were originally developed in the context of single-level regression; however, they have been applied without modification to multilevel settings (e.g., Clark, 2013; Kreft & De Leeuw, 1998; Shieh & Fouladi, 2003; Yu et al., 2015).

In single-level regression, the impacts of collinearity are well understood. Its primary consequence is large standard errors (*SE*s) of the slope estimates associated with the collinear predictors. If a predictor is highly collinear with other predictor(s), the variance (and accordingly, *SE*) of its slope estimate will increase; we become less able to distinguish its unique contribution to explaining variance in the outcome, and therefore its contribution is estimated with less precision (Farrar & Glauber, 1967; Willan & Watts, 1978). A related consequence is that the slope estimates associated with collinear predictors will be highly variable from sample to sample and often assume substantively impossible values. Additionally, large *SE*s result in reduced power to detect effects (Dormann et al., 2013; O'Brien, 2007).

Methodologists have debated the extent to which it is useful to investigate, diagnose, and attempt to remedy collinearity problems. First, in single-level regression, although collinearity causes large *SE*s, these *SE*s are not biased, which has informed arguments that we need not consider them problematic (Goldberger, 1991). Second, collinearity is just one of many possible causes of large *SE*s, which has led some methodologists to argue that the dangers of collinearity are often exaggerated (e.g., Mason & Perreault, 1991; O'Brien, 2007). Other factors, including sample size, variances of the predictors and outcome, and the predictors' explanatory power (R_y^2), play equally important roles in determining the *SE*s of regression estimates. Rather than blame exclusively collinearity for large *SE*s, it is necessary to consider its interactions with these other data and model characteristics (Hayo, 2018). Third, it is often argued that observed collinearity reflects the natural state of the system under study, and therefore, attempts to remedy it are futile (Goldberger, 1991; however, collinearity in a data set does not always reflect collinearity in the population, such as due to range restriction in the sample). On the other hand, investigating collinearity often yields actionable insight. It may spur the researcher to consider whether redundant variables are each theoretically important, leading to a more parsimonious model. We can contrast this with the other factors that contribute to *SE*s, such as R_y^2, that are largely outside the researcher's control. Most importantly, arguments for and against the relevance of collinearity are rooted in knowledge of single-level, ordinary least squares (OLS) regression. Due to very limited knowledge of collinearity's effects on multilevel models, it is unclear whether similar arguments could be made in multilevel settings. To inform a balanced discussion of whether and how collinearity should be diagnosed and remedied in multilevel settings, we must understand *how* collinearity impacts multilevel models in the first place.

The effects of collinearity are most relevant when the purpose of analysis is to make inferences about the estimated regression coefficients. This may not always be the central goal of a statistical model. For example, in machine learning applications, models may include many predictors whose individual slopes are not of concern in terms of significance or meaning; rather, the whole model's ability to predict *y* is optimized. In other applications, it is critically important to examine and interpret each estimated regression coefficient; its statistical significance, point estimate, and confidence interval will be used to inform later hypotheses and theory. The latter scenario is more typical of social science research, and therefore, throughout this study we assume the primary goal is to make inferences about regression coefficients.

Collinearity in Multilevel Models

In this study, we focus our attention to multilevel modeling applications in which it is substantively important to model relations both within and across units of clustering. For example, in educational studies of students nested within classrooms or schools, it is often essential to disaggregate the effects of key variables; this is useful for identifying "effective schools" in terms of average student outcomes (Raudenbush & Willms, 1995) or for isolating the effects of an intervention at either the student or the classroom level. This is not always the case for multilevel data sets; in some applications, nesting structure is treated as a nuisance factor and it is not substantively relevant to model level-specific relations. Here we assume that it is theoretically important to consider the hierarchical structure of the data, and as such, it is essential to consider level-specific relations separately (Snijders & Bosker, 2012).

The consequences of collinearity have been scarcely studied in the context of multilevel models. Empirical observations were the first to suggest that collinearity may have similar consequences in multilevel models as it does in single-level regression, namely, large *SE*s and unstable point estimates (Bonate, 1999; Kreft & de Leeuw, 1998; Kubitscheck & Hallinan, 1999). Since then, a few simulation studies have more systematically investigated how collinearity impacts multilevel models. Typically, outcomes of interest are point estimates and their relative bias (for both fixed effects and random effect (co)variances) and *SE*s of the fixed effect estimates.

Regarding collinearity's consequences for *SE*s, findings indicate that as predictor correlation increases, *SE*s of fixed effect estimates also increase. Some studies report increases in *SE*s without commenting on their bias (Clark, 2013; Yu et al., 2015). Others report that *SE*s become upwardly biased as predictor collinearity increases (Blaze & Ye, 2012; Shieh & Fouladi, 2003); estimated *SE*s were larger than the true standard deviation of the fixed effect estimates across iterations. In multilevel models it remains unclear to what degree large *SE*s are due to a true increase in the sampling variability of the estimates, estimation bias, or both.

Findings with respect to random effects are less consistent. In a standalone empirical example, Hendrickx (2018) suggested that collinearity may cause greater instability in random effect (co)variance estimates than fixed effect estimates. Two simulation studies have reported outcomes regarding the bias of variance estimates, but their results are contradictory. Shieh and Fouladi (2003) observed that as predictor correlation increased at level 1, most variance estimates exhibited small-to-moderate downward bias. In contrast, Yu et al. (2015) reported no bias in any variance estimates across all collinearity conditions at level 1 or level 2. They attributed this result to their use of restricted maximum likelihood (REML) estimation (Shieh & Fouladi used iterative generalized least squares); however, this result has yet to be replicated. Whether collinearity leads to biased estimates of variance components remains an unresolved question. The current study will address this issue.

Simulations have begun to probe the conditions under which the consequences of collinearity may be exacerbated or mitigated in multilevel data. The design factors most commonly varied are cluster size, number of clusters, and the intraclass correlation of the outcome (ICC_y).

Findings with respect to sample size are consistent: all else being equal, as cluster size and number of clusters increases, the harmful effects of collinearity are mitigated. Findings regarding ICC_y remain inconclusive. Shieh and Fouladi (2003) found no effect of ICC_y, whereas other studies (Blaze & Ye, 2012; Yu et al., 2015) observed that, all else being equal, the effects of collinearity were mitigated as ICC_y increased. The effects of many other important design factors remain unknown. For example, there is speculation that strong correlations among the random effects may be important to consider alongside observed predictor collinearity (Stinnett, 1994; Zhang & Chen, 2013), however this has never been systematically examined. Additionally, Mela and Kopalle (2002) showed that the sign of the predictor correlation can exert asymmetric effects on bias and *SE*s in single-level regression, but direction of predictor collinearity has never been studied in the context of multilevel models.

Across simulation studies, findings regarding the bias of fixed effect estimates have been consistent; fixed effect estimates remain unbiased despite any collinearity present at either level. Such results are consistent with the single-level regression case. The current literature suggests that this finding is conclusive; however, major limitations of past work – specifically, a lack of attention to centering and disaggregation across levels – call this conclusion into question.

Collinearity and Centering in Multilevel Models

In the broader multilevel modeling literature, the importance of centering and level-specificity has been discussed at length. Methodologists often advocate for disaggregating predictors into their level-specific (i.e., level 1 and level 2) components (Cronbach & Webb, 1975; Curran & Bauer, 2010; Hofmann & Gavin, 1998; Kenny & La Voie, 1985; Robinson, 1950). This is typically accomplished by centering level-1 predictors around their cluster means (as cluster mean centering isolates the "pure" level 1 component of the predictor) and including cluster means themselves as predictors at level 2 (as cluster means are the "pure" level 2 component of the predictor). Under disaggregation, slopes are also level-specific and can be interpreted as within- and between-cluster effects. When uncentered predictors are used in isolation (when disaggregation is not conducted), this yields "conflated" slope estimates that are uninterpretable mixes of within- and between-cluster slopes (Enders & Tofighi, 2007; Raudenbush & Bryk, 2002; Yaremych et al., 2021). Conflated slope estimates are systematically affected by many arbitrary factors, including sample size at each level, ICC_y, and predictor ICCs (Lüdtke et al., 2008; Raudenbush & Bryk, 2002).

The importance of centering has received little attention in the multilevel literature concerning collinearity. Most simulation studies have either briefly mentioned or not discussed centering, often making it unclear whether uncentered or level-disaggregated predictors were under study. This is concerning because it is well-documented that centering choice has direct consequences for parameter estimates and their substantive interpretations. Thus, we can expect that the consequences of collinearity will differ in important ways according to centering choice. Additionally, contradictory methodological recommendations highlight the need for clarification of how centering choice may impact the nature of collinearity itself. Hendrickx (2018) argued that collinearity should be examined among the "untransformed" variables in multilevel settings and discouraged any type of centering, whereas Bickel (2007) suggested grand mean centering as the best approach to reduce collinearity problems in multilevel models, borrowing arguments from single-level settings without modification. How collinearity and its consequences may change according to centering choice is an open question that has yet to be addressed.

Centering and disaggregation also have been ignored in discussions of *diagnosing* collinearity in multilevel data. In multilevel data, collinearity can arise both within and between clusters; therefore, we can expect that successfully identifying collinearity problems will hinge on whether they are diagnosed accurately at each level. However, no extant work has discussed the importance of level-specificity in collinearity diagnosis. Diagnostics developed in single-level settings, such as the Variance Inflation Factor (VIF) and condition number (κ), have been applied to multilevel data without modification or consideration of centering.

Study Goals

The purpose of this study is to integrate the topics of centering and collinearity in multilevel models and develop novel insights into their interactivity. To this end, the study has three primary goals. First, we demonstrate how different centering choices for level-1 predictors (i.e., the use of uncentered vs. level-disaggregated predictors) lead to different impacts of collinearity in terms of biased point estimation, magnitudes of *SE*s, and biased *SE*s. Based on aforementioned findings in the centering literature, we expect that predictor collinearity will be yet another data characteristic that systematically affects conflated slope estimates. In contrast, we hypothesize that level-specific slope estimates will *not* be affected by collinearity alone. In other words, we anticipate that point estimates will vary as a function of predictor collinearity in the conflated model, but not in the disaggregated model. We will also explore how collinearity affects point estimates in a *partially* disaggregated model, where some level-1 predictors are disaggregated whereas others are not. Second, we clarify how other data characteristics, including ICC_y, predictor ICCs, and correlations among the random effects, may exacerbate or mitigate the effects of predictor collinearity across different centering specifications. Third, we demonstrate the importance of disaggregation for the diagnosis of collinearity in multilevel data and provide recommendations for diagnosing collinearity in a level-specific manner.

Analytic Developments

The Population Model

To preface our discussion of within- and between-cluster correlations and effects, we begin with a population model similar to that presented by Snijders and Bosker (2012). In this model we consider a level-1 outcome variable y_{ij} and a pair of level-1 predictors, x_{1ij} and x_{2ij}, each of which is a level-1 observation *i* within cluster *j*. For each variable, there is a population mean (denoted μ_y, μ_{x_1}, μ_{x_2}), a latent main effect of cluster *j* (denoted U_{yj}, $U_{x_{1j}}$, $U_{x_{2j}}$), and an individual-level deviation from that effect (denoted e_{yij}, $e_{x_{1ij}}$, $e_{x_{2ij}}$). Individual-level deviations are assumed to follow a normal distribution with mean zero, e.g., N(0, σ_y^2).

$$y_{ij} = \mu_y + U_{yj} + e_{yij};\;\; x_{1ij} = \mu_{x_1} + U_{x_{1j}} + e_{x_{1ij}};\;\; x_{2ij} = \mu_{x_2} + U_{x_{2j}} + e_{x_{2ij}} \tag{1}$$

The within-cluster effects of the predictors are denoted β_{W1}, β_{W2}, and are obtained from regressing y_{ij} on within-group deviations from each cluster mean:

$$y_{ij} = \mu_y + U_{yj} + \beta_{W1}\left(x_{1ij} - \mu_{x_1} - U_{x_{1j}}\right) + \beta_{W2}\left(x_{2ij} - \mu_{x_2} - U_{x_{2j}}\right) + e_{yij} \tag{2}$$

The population between-cluster effects of the predictors are denoted β_{B1}, β_{B2}. These effects are obtained by regressing the grand-mean-centered population group mean of y_{ij} on the grand-mean-centered population group means of the predictors. Population group means are latent (i.e., unobservable), and therefore, population between-cluster effects are also latent. Latent means are assumed to follow a multivariate normal distribution.

$U_{yj} = \beta_{B1}U_{x_{1j}} + \beta_{B2}U_{x_{2j}} + e_{yj}$, where e_{yj} is a group-level residual. (3)

Between-cluster effects can also be estimated from observed group means, which are denoted by $\bar{y}_j$, $\bar{x}_{1j}$, and $\bar{x}_{2j}$. Latent group means are not the same as observed group means, as observed group means inherently contain measurement error (Snijders & Bosker, 2012). A population (i.e., latent) group mean is related to an observed group mean by the reliability of the observed group mean; when the reliability of the observed group mean is 1, the two means are equal. When we work with sample data, population means are replaced with sample means (i.e., observed group means). The models presented later in this paper therefore involve observed group means.

Note that level-specific effects (i.e., within- and between-cluster effects) are obtained by disaggregating the level-1 predictor into its level-specific components. The cluster mean centered predictor (its level-1 component) yields the within-cluster effect, whereas the cluster mean of the predictor (its level-2 component) yields the between-cluster effect.

Multilevel Covariance and Correlation

We next discuss covariance and correlation at the between-cluster level, the within-cluster level, and conflated across levels (termed *total correlations* and *total covariance*) in multilevel data. Because collinearity is often defined in terms of predictor correlations, it is important to demonstrate how total and level-specific correlations arise and how they are related to one another. To do so, we integrate two lines of work that have addressed this issue separately.

Building upon the original work of Robinson (1950), Gale (1987) and Snijders and Bosker (2012) presented expressions for the correlation of level-1 variables, x_{1ij} and x_{2ij}, at each level of analysis. For a given cluster j, we define the following sums of squares:

$$SS_{x_{1j}} = \sum_{i=1}^{n_j} \left(x_{1ij} - \bar{x}_{1j} \right)^2 \quad ; \quad SS_{x_{2j}} = \sum_{i=1}^{n_j} \left(x_{2ij} - \bar{x}_{2j} \right)^2 \qquad (4)$$

For a given cluster j, we can compute the within-cluster correlation using each observation's deviation from the cluster mean:

$$r_j = \sum_{i=1}^{n_j} \left(x_{1ij} - \bar{x}_{1j} \right)\left(x_{2ij} - \bar{x}_{2j} \right) \Big/ \left(SS_{x_{1j}} \right)^{1/2} \left(SS_{x_{2j}} \right)^{1/2} \qquad (5)$$

The overall within-cluster correlation, r_W, is then a weighted average of the within-cluster correlations from clusters $j = 1, 2, \ldots J$:

$$r_W = \sum_{j=1}^{J} r_j \left(SS_{x_{1j}}\right)^{1/2} \left(SS_{x_{2j}}\right)^{1/2} \Bigg/ \left(\sum_{j=1}^{J} SS_{x_{1j}}\right)^{1/2} \left(\sum_{j=1}^{J} SS_{x_{2j}}\right)^{1/2} \tag{6}$$

Similarly to between-cluster regression effects, the between-cluster correlation, r_B, is computed differently if population (i.e., latent) group means or observed group means are used. If the reliabilities of the group means are 1, then these correlations are equivalent. Here, for simplicity, we assume this is the case[1]. Therefore, r_B is given by the following equation, where $\bar{x}_1$ and $\bar{x}_2$ denote the observed grand mean of x_{1ij} and x_{2ij}, respectively:

$$r_B = \sum_{j=1}^{J} n_j \left(\bar{x}_{1j} - \bar{x}_1\right)\left(\bar{x}_{2j} - \bar{x}_2\right) \Bigg/ \left(\sum_{j=1}^{J} n_j \left(\bar{x}_{1j} - \bar{x}_1\right)^2\right)^{1/2} \left(\sum_{j=1}^{J} n_j \left(\bar{x}_{2j} - \bar{x}_2\right)^2\right)^{1/2} \tag{7}$$

The total correlation, r_T, is a combination of r_W and r_B. Importantly, r_T is the correlation that arises among uncentered level-1 variables. r_T is given by:

$$r_T = \sqrt{\text{ICC}_{x_{1ij}}}\sqrt{\text{ICC}_{x_{2ij}}}\, r_B + \sqrt{1-\text{ICC}_{x_{1ij}}}\sqrt{1-\text{ICC}_{x_{2ij}}}\, r_W \tag{8}$$

Where $\text{ICC}_{x_{1ij}}$ and $\text{ICC}_{x_{2ij}}$ are the intraclass correlations of x_{1ij} and x_{2ij}, respectively. This formula pertains to only one total correlation at a time and has not been generalized to a full correlation matrix, $\mathbf{R_T}$.

Second, Muthén (1990) demonstrated how the total covariance matrix for level-1 variables in multilevel data, $\boldsymbol{\Sigma}_T$, is related to the between-cluster covariance matrix, $\boldsymbol{\Sigma}_B$, and the within-cluster covariance matrix, Σ_W, via a simple sum:

[1] See Snijders and Bosker (2012), section 3.6.2, for greater detail.

$$\boldsymbol{\Sigma}_T = \boldsymbol{\Sigma}_B + \boldsymbol{\Sigma}_W. \tag{9}$$

In the sample this is denoted $\mathbf{S}_T = \mathbf{S}_B + \mathbf{S}_W$.

In Appendix A, we standardize Muthén's total covariance matrix $\boldsymbol{\Sigma}_T$ to yield a total correlation matrix, $\mathbf{R_T}$, and show that the off-diagonal elements of this matrix (i.e., the total correlation coefficients) are equal to those derived for a single correlation coefficient (Gale, 1987; Snijders & Bosker, 2012). In other words, the formula for a total correlation coefficient agrees with pairwise correlations obtained from standardizing Muthén's total covariance matrix. We also show that a total (i.e., conflated) correlation coefficient can be expressed in two equivalent ways. The first is Eq. 8, and the second is:

$$r_T = \frac{\text{cov}(x_{1ij}, x_{2ij})_B}{SD_{x_{1ij}} SD_{x_{2ij}}} + \frac{\text{cov}(x_{1ij}, x_{2ij})_W}{SD_{x_{1ij}} SD_{x_{2ij}}} \tag{10}$$

Where $\text{cov}(x_{1ij}, x_{2ij})_B$ arises from cluster means' deviations around the grand mean, and $\text{cov}(x_{1ij}, x_{2ij})_W$ arises from individual observations' deviations around their respective cluster means. Eqs. 8 and 10 are equivalent.

Both representations of the conflated correlation coefficient can aid our understanding of the factors that determine its value. Beginning with Eq. 8, r_T can be understood as a weighted combination of r_B and r_W. The between-cluster correlation, r_B, is weighted by $\sqrt{\text{ICC}_{x_{1ij}}}\sqrt{\text{ICC}_{x_{2ij}}}$, indicating that r_B receives more weight as the ICC of each variable increases, and as ICCs of both variables approach 1, r_T approaches r_B. Similarly, r_W is weighted by $\sqrt{1-\text{ICC}_{x_{1ij}}}\sqrt{1-\text{ICC}_{x_{2ij}}}$, indicating that r_W receives more weight as the ICC of each variable decreases, and as ICCs of both variables approach 0, r_T approaches r_W. Next, Eq. 10 suggests the total dependence between x_{1ij} and x_{2ij} can be understood as the sum of two parts: the proportion of that dependence that occurs between clusters (i.e., between-cluster covariance divided by total dispersion of each predictor), and the proportion of that dependence that occurs within clusters (i.e., within-cluster covariance divided by total dispersion of each predictor).

Eqs. 8 and 10 also lend insight into the situations in which r_T is meaningful or meaningless. It becomes clear that $r_T = r_B$ when the ICC of each variable is 1, and $r_T = r_W$ when the ICC of each variable is 0. Otherwise, even if $r_B = r_W$, r_T will be equal to neither level-specific correlation. For example, let $r_B = r_W = .5$. Suppose that most of x_{1ij}'s variance is within clusters, such that $ICC_{x_{1ij}} = .03$, whereas most of x_{2ij}'s variance is between clusters, such that $ICC_{x_{2ij}} = .60$. Then, according to Eq. 8, $r_T = .37$. The total correlation is misleading, as are any conclusions informed by it, even if level-specific correlations are equal.

Thus, in most cases, r_T will be an uninterpretable weighted combination of level-specific correlations. In many situations, even if each level-specific correlation is strong, r_T may suggest that there is no significant relation among the variables. Additionally, when r_B and r_W are the same sign, r_T need not lie between them and may be closer to zero than both (Gale, 1987; Snijders & Bosker, 2012). See Online Appendix A for details about the conditions that yield this pattern. Overall, the lack of interpretability of the conflated correlation is exposed here.

Consequences of Collinearity for Multilevel Model Estimation: $\boldsymbol{\beta}_{GLS}$ Derivation

A primary goal of this study is to demonstrate analytically how collinearity's consequences change across different centering specifications. Here, we focus on the conflated model, wherein uncentered level-1 predictors are used in isolation. Past work mathematically shows conflated slope estimates to be affected by a variety of extraneous data characteristics; here we show that collinearity among predictors (operationalized as covariance) is yet another data characteristic to systematically affect these estimates. Currently, most insight into how extraneous data characteristics influence conflated slope estimates comes from a classic derivation by Raudenbush and Willms (1995). Assuming a balanced design, they show:

$$\gamma_{10}^{*} = \frac{W_1\hat{\beta}_B + W_2\hat{\beta}_W}{W_1 + W_2}; \quad W_1 = \left[\mathrm{var}(\hat{\beta}_B)\right]^{-1}; \quad W_2 = \left[\mathrm{var}(\hat{\beta}_W)\right]^{-1} \tag{11}$$

Where $\hat{\beta}_B$ is the estimated between-cluster effect and $\hat{\beta}_W$ is the estimated within-cluster effect. The conflated slope estimate is a weighted average of level-specific effects, where the weights reflect the precision of each estimate. It follows that factors influencing the precision of each level-specific estimate (e.g., sample size, ICC_y) systematically affect the conflated slope estimate. Eq. 11 pertains to a single slope estimate rather than those of multiple, potentially collinear, predictors. A more general expression is needed to accommodate any number of predictors and their potential interrelations. We derive that expression here.

Multilevel models are typically estimated via maximum likelihood (ML), but ML estimates are algebraically intractable. The generalized least squares (GLS) estimator is algebraically tractable and yields conflated slope estimates that are asymptotically equivalent to ML estimates (Raudenbush & Bryk, 2002). Therefore, in this section we employ the GLS estimator as the basis for our analysis. The estimator was first introduced because it is more efficient than OLS estimation when $\beta_B = \beta_W$. Scott and Holt (1982) defined the estimator as follows:

$$\hat{\beta}_{GLS} = \left\{ \sum_{j=1}^{J} \frac{X_{Bj}^T X_{Bj}}{1+(n_j-1)\rho} + \sum_{j=1}^{J} \frac{X_{Wj}^T X_{Wj}}{1-\rho} \right\}^{-1} \times \left\{ \sum_{j=1}^{J} \frac{X_{Bj}^T Y_j}{1+(n_j-1)\rho} + \sum_{j=1}^{J} \frac{X_{Wj}^T Y_j}{1-\rho} \right\}, \tag{12}$$

where J is the number of clusters, n_j is cluster size for cluster j, ρ is ICC_y, X_{Bj} is a vector of cluster means for cluster j, and X_{Wj} is a vector of cluster mean centered predictors for cluster j.

Eq. 12 illuminates how a conflated slope estimate for a single predictor can be derived. The authors note that the formula could be scaled up to multiple predictors; however, this has never been shown. After adapting Scott and Holt's (1982) notation to conform to notation now in common use (e.g., Snijders & Bosker, 2012), we derived the maximally general form of the GLS estimator, which allows for any number of predictors, any number of clusters, and potentially variable cluster size. (See Appendix B for details of the derivation):

$$\hat{\beta}_{GLS} = \left\{ \sum_{j=1}^{J} \left[\begin{array}{c|c} \left(1+\left(n_j-1\right)\rho\right)^{-1} n_j & \left(1+\left(n_j-1\right)\rho\right)^{-1} n_j \bar{\mathbf{x}}_j' \\ \hline \left(1+\left(n_j-1\right)\rho\right)^{-1} n_j \bar{\mathbf{x}}_j & \left(1+\left(n_j-1\right)\rho\right)^{-1} n_j \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j' + \left(1-\rho\right)^{-1} \left(\mathbf{X}_j'\mathbf{X}_j - n_j \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j'\right) \end{array} \right] \right\}^{-1} \times \left\{ \sum_{j=1}^{J} \left(\left(1+(n_j-1)\rho\right)^{-1} \left[\mathbf{1}_{n_j} \mid \mathbf{1}_{n_j} \bar{\mathbf{x}}_j' \right]^T Y_j + \left(1-\rho\right)^{-1} \left[\mathbf{0}_{n_j} \mid \mathbf{X}_j - \mathbf{1}_{n_j} \bar{\mathbf{x}}_j' \right]^T Y_j \right) \right\} \quad (13)$$

where $\mathbf{1}_{n_j}$ is a $n_j \times 1$ column vector of 1's, $\mathbf{0}_{n_j}$ is a $n_j \times 1$ column vector of 0's, $\bar{\mathbf{x}}_j$ is a column vector of cluster means in cluster j whose first element is 1 as a multiplier for the intercept, $\mathbf{X}_j$ is the original data matrix in cluster j whose first column is 1 as a multiplier for the intercept, and Y_j is a column vector of outcomes in cluster j. This derivation transforms Scott and Holt's (1982) original formula from scalar notation to matrix notation. Thus, we show how a vector of multiple conflated slope estimates ($\hat{\beta}_{GLS}$) can arise.

From the maximally general form of the GLS estimator, we next show that conflated slope estimates are systematically biased by covariance (i.e., collinearity) among predictors at each level. To do so, we demonstrate that Eq. 13 can be expressed in terms of between- and within-cluster covariance matrices of the predictors. Equations for the sample between-cluster covariance matrix, $\mathbf{S}_B$, and the sample pooled within-cluster covariance matrix, $\mathbf{S}_{PW}$, were introduced by Muthén (1990) as:

$$\mathbf{S}_B = (J-1)^{-1} \sum_{j=1}^{J} n_j \left(\bar{x}_{.j} - \bar{x}_{..}\right)\left(\bar{x}_{.j} - \bar{x}_{..}\right)' \quad (14)$$

$$\mathbf{S}_{PW} = (N-J)^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n_j}\left(x_{ij}-\overline{x}_{.j}\right)\left(x_{ij}-\overline{x}_{.j}\right)' \tag{15}$$

where N is total sample size, x_{ij} is a level-1 observation, $\overline{x}_{.j}$ is a cluster mean in cluster j, and $\overline{x}_{..}$ is the grand mean.

In Appendix C, after assuming equal cluster sizes for simplicity, we show that a key between-cluster component of Eq. 13 can be expressed in terms of Muthén's $\mathbf{S}_B$:

$$\sum_{j=1}^{J}\left(1+\left(n_j-1\right)\rho\right)^{-1} n_j\overline{\mathbf{x}}_j\overline{\mathbf{x}}_j' = \left(1+\left(n-1\right)\rho\right)^{-1} n\begin{bmatrix} J & \sum_{j=1}^{J}\overline{\mathbf{z}}_j' \\ \sum_{j=1}^{J}\overline{\mathbf{z}}_j & (J-1)n^{-1}\mathbf{S}_B \end{bmatrix} \tag{16}$$

And we show the within-cluster component of Eq. 13 can be expressed in terms of $\mathbf{S}_{PW}$:

$$\sum_{j=1}^{J}(1-\rho)^{-1}\left(\mathbf{X}_j'\mathbf{X}_j - n_j\overline{\mathbf{x}}_j\overline{\mathbf{x}}_j'\right) = \left(1-\rho\right)^{-1}\begin{bmatrix} 0 & \sum_{j=1}^{J}\left(\mathbf{Z}_j - n\overline{\mathbf{z}}_j'\right) \\ \sum_{j=1}^{J}\left(\mathbf{Z}_j' - n\overline{\mathbf{z}}_j\right) & \left(N-J\right)\mathbf{S}_{PW} \end{bmatrix} \tag{17}$$

Where $\overline{\mathbf{z}}_j$ is a vector of cluster means in cluster j, excluding the 1 in the first element, and $\boldsymbol{Z}_j$ is the original data matrix in cluster j, excluding the first column of 1's. In summary, both the between- and within-cluster covariance matrices of predictors appear in the maximally general form of the GLS estimator. Therefore, we have proven that predictor covariance (i.e., collinearity) directly informs conflated slope estimates.

Effect of Collinearity at Each Level

As demonstrated above, the maximally general form of the GLS estimator can be expressed in terms of the between- and within-cluster covariance matrices of predictors. Interestingly, only level-specific (co)variances appear in Eqs. 16 and 17. This indicates that even if level-specific slopes are not being estimated, level-specific predictor collinearity is a factor that systematically "pushes around" point estimates. Therefore, in all cases it will be essential to diagnose collinearity in a level-specific manner.

In single-level settings, properties of the $\mathbf{X'X}$ matrix have been exploited to show how the *sampling variances* of parameter estimates are enlarged by collinearity. In OLS regression, collinearity does not bias point estimation. We have shown that in multilevel models, this property does not always apply. When uncentered predictors are used in isolation in the multilevel model, predictor collinearity plays a role in determining the point estimates themselves – not just their precisions.

Effects of Other Data Characteristics

Our derivations illuminate how other data characteristics will interact with predictor covariance to either exacerbate or mitigate the effects of collinearity at each level. We begin with ICC_y (denoted with ρ in Eqs. 13, 16, and 17). Accounting for all weights in Eq. 17, $\mathbf{S}_{PW}$ is weighted by $(N-J)(1-\rho)^{-1}$. This term increases as ICC_y increases. However, Eq. 13 indicates that the matrix containing this term is ultimately inverted. Accounting for all inversions, as ICC_y increases, within-cluster covariance is given less weight. Therefore, we expect that any harmful effects of collinearity at the within-cluster level will be mitigated by a higher ICC_y and exacerbated by a lower ICC_y.

Between-cluster covariance is weighted by a more complex term. Simplifying Eq. 16, $\mathbf{S}_B$ is weighted by $(J-1)(1+(n-1)\rho)^{-1}$. This term decreases as ICC_y increases. However, again, the matrix containing this term is inverted in Eq. 13. After accounting for all inversions, the term involving between-cluster covariance is given more weight as ICC_y increases. As a result, we expect that the harmful effects of collinearity at the between-cluster level will be exacerbated by higher ICC_y. However, because this weight is not purely informed by ICC_y and involves other important quantities, the influence of ICC_y may not be as direct as that at the within-cluster level.

Finally, sample sizes at both levels appear in the GLS estimator. $\mathbf{S}_{PW}$ is weighted by a term involving total sample size (*N*) and number of clusters (*J*), whereas the weight associated with $\mathbf{S}_B$ contains number of clusters (*J*) and cluster size (*n*). All these terms increase as sample size increases. However, accounting for the fact that these terms are ultimately inverted, both within- and between-cluster covariance receive less weight as sample size increaes. As expected, larger sample size will mitigate the harmful effects of collinearity at either level in multilevel data.

Importance of Disaggregation for Collinearity Diagnostics

In addition to demonstrating that collinearity biases fixed effect estimates when predictors are uncentered, we can also demonstrate the consquences of attempting to diagnose collinearity on uncentered predictors. Here we combine what we have shown regarding multilevel covariance and correlation, together with the mathematics underlying common collinearity diagnostics, to gain insight into diagnostics' behavior when applied to uncentered predictors vs. predictors that have been disaggregated into level-specific components.

Variance Inflation Factors (VIFs)

In single-level regression, perhaps the most popular collinearity diagnostic is the VIF. A VIF is computed for each predictor x_i, and is defined as the multiplicative factor by which the variance of the regression coefficient associated with that predictor, $var(\hat{\beta}_i)$, has increased due to collinearity in the sample. VIFs can be calculated in two equivalent ways. The first is to regress x_i on all other predictors in the model, find the multiple $R^2_{x_i}$ from that regression, and compute $1/(1 - R^2_{x_i})$. The second is to obtain the correlation matrix of predictors, $\mathbf{R}$, invert it to obtain $\mathbf{R}^{-1}$, and extract the diagonal elements of $\mathbf{R}^{-1}$, which are the VIF for each predictor.

Given this knowledge, we can draw conclusions about VIFs' behavior in multilevel data when predictors are conflated vs. disaggregated across levels.

The centering method chosen for the predictors directly informs their correlation matrix and the resulting VIFs. Disaggregation results in level-specific sets of predictors that are uncorrelated across levels; VIFs are informed by $\mathbf{R_W}$ and $\mathbf{R_B}$, yielding two sets of VIFs that are level-specific and mutually independent. However, if uncentered predictors are used, VIFs are informed by $\mathbf{R_T}$. The correlations within $\mathbf{R_T}$ are weighted combinations of between- and within-cluster correlations and are almost always uninterpretable. As a result, collinearity diagnostics informed by $\mathbf{R_T}$ are also uninterpretable and misleading. In situations where $\mathbf{R_T}$ suggests weak relationships even though level-specific correlations are strong, VIFs will not indicate collinearity problems, and those problems will go undetected. Thus, we cannot assert that disaggregating predictors *necessarily reduces* collinearity in the data or its associated VIFs. Rather, disaggregation enables identification of level-specific collinearity problems that are meaningful and relevant to estimation. Level-specific collinearity is a determinant of point estimates in all cases (even in the conflated model), so it is essential to diagnose collinearity in a level-specific manner in order to understand how parameter estimates and their precision have been impacted.

Condition Number (κ)

Another popular collinearity diagnostic is the condition number (κ). κ is predicated on the fact that, given an eigen-decomposition of a symmetric matrix, eigenvalues of zero indicate exact linear dependencies within the matrix**.** Eigenvalues *near* zero indicate *near* linear dependencies. κ is computed via an eigen-decomposition of $\mathbf{X'X}$ (where $\mathbf{X}$ is the observed predictor set), yielding the eigenvalues $\lambda_{min}, \lambda_2, \lambda_3, \ldots, \lambda_{max}$. κ is given by: $\kappa = \sqrt{\lambda_{max}/\lambda_{min}}$. Greater dissimilarity of λ_{max} and λ_{min} indicates greater linear dependency within the data matrix, so larger κ indicates stronger collinearity. κ is calculated from standardized data (i.e., all variables are scaled to have unit variances), so variable scales are irrelevant.

For eigenvalue-based diagnostics, Belsley (1991) demonstrated the benefits of orthogonal (or nearly orthogonal) subsets of predictors: when one set of predictors is orthogonal to a second set, any collinearity present in the first set will have no effect on the collinearity diagnostics or *SE*s associated with the second set. This property is directly relevant to multilevel data, where disaggregation results in level-specific sets of predictors that are orthogonal. If predictors are disaggregated into their level-specific parts, then any collinearity present at a given level has no impact on the diagnosis or consequences of collinearity in the other level.

Past researchers (e.g., Yu et al., 2015) have recommended calculating an "overall" κ for multilevel data, using all level-1 and level-2 predictors. We argue that this "overall" κ lacks utility. Recall that κ is derived from λ_{max} and λ_{min}. Consider the scenario where the predictor set consists entirely of level-1 predictors that have been disaggregated into their level-specific parts. Because κ is applied to standardized variables, the sum of eigenvalues for the level-1 and level-2 predictor sets will be equal. As more collinearity is present at a given level, λ_{max} and λ_{min} are driven upward and downward, respectively, such that the sum remains the same. Therefore, the level containing greater collinearity will yield both the largest and smallest eigenvalues. The "overall" κ will be entirely driven by whichever level has the stronger collinearity. Alternatively, there may be an unequal number of predictors at each level, so the sum of the eigenvalues at each level is not equal. In this case, λ_{max} and λ_{min} could come from different levels, creating an even more meaningless κ. In summary, an overall κ diagnostic will at best be insufficient for identifying level-specific collinearity, and will at worst be misleading, suggesting that separate level-specific κs will be much more useful.

Demonstration on Simulated Examples

The above conclusions can be demonstrated with simple simulated examples. We generated multilevel data sets with three level-1 predictors, x_{1ij}, x_{2ij}, and x_{3ij}, which had various level-specific correlation structures. In all conditions, x_{3ij} was minimally correlated with the other predictors. Correlations between the first two predictors were the focus of the manipulation, which we will refer to as r_W, r_B, and r_T in this section. The data generating model for Condition 1 was $r_W = r_B = 0$ and $\text{ICC}_{x_{1ij}} = \text{ICC}_{x_{2ij}} = \text{ICC}_{x_{3ij}} = .25$. Condition 1 was created as a baseline case with no collinearity and typical ICCs. In Condition 2, $r_W = .7$, $r_B = -.9$, and $\text{ICC}_{x_{1ij}} = \text{ICC}_{x_{2ij}} = \text{ICC}_{x_{3ij}} = .25$. Condition 2 was chosen to demonstrate the consequences of conflation in situations where level-specific correlations are highly dissimilar. In Condition 3, $r_W = .25$, $r_B = .95$, and ICC was manipulated such that $\text{ICC}_{x_{1ij}} = .8$, $\text{ICC}_{x_{2ij}} = .01$, and $\text{ICC}_{x_{3ij}} = .25$. Condition 3 was chosen because these data characteristics give rise to a conflated correlation that is smaller than both the within- and between-cluster correlations. (For details about the data conditions that give rise to $r_T < r_W$ and $r_T < r_B$, see Online Appendix A.)

Having generated each data set, we computed correlations, VIFs, and κs using both the conflated and disaggregated predictor set. When calculating κ, we computed an "overall" measure for all predictors, and then separately for the level-1 set and the level-2 set. See Table 1.

From the correlation matrices of our simulated data, we see that disaggregating the predictors introduces orthogonality across the level-specific components; all level-1 components have zero correlation with all level-2 components. Additionally, uncentered predictors yield correlations that are unequal to any level-specific correlations; in conditions 1 and 2, r_T is equal to neither r_W nor r_B and falls between them, and in condition 3, r_T is notably smaller than both r_W and r_B.

The necessity of disaggregation also becomes clear in the collinearity diagnostics. For example, in condition 2, each level-specific correlation is strong, but r_T falls between them and is closer to zero than both. VIFs associated with uncentered predictors falsely suggest little-to-no collinearity. In contrast, for the disaggregated predictors, VIFs correctly identify strong level-specific collineary. The same problem occurs in condition 3 even though r_W and r_B are not of opposite sign. Due to dissimilar ICCs of the predictors, $\mathbf{R_T}$ yields misleadingly small VIFs, even though predictor collinearity is strong at both levels, particularly at level 2.

The consequences of conflation are the same for κs as they are for VIFs. Additionally, for disaggregated predictors, the overall κ (computed for the entire predictor set) is always equal to the level-specific κ from the level with stronger collinearity. For example, in condition 3, overall κ is equal to κ obtained from only the level-2 variables. From the overall measure it is impossible to ascertain whether collinearity is problematic at level 1, level 2, or both. Computing separate level-specific κs is much more useful for identifying meaningful collinearity problems.

Summary

In this section, we have analytically demonstrated that level-specific covariance (i.e., collinearity) is a factor that determines point estimates for conflated slopes. This is in direct contrast to single-level regression where collinearity never biases point estimation. The β_{GLS} derivation indicates how other data characteristics, particularly ICC_y and cluster size, will interact with collinearity at each level to either exacerbate or mitigate its effects. We have also demonstrated that diagnostics applied to uncentered predictors are misleading, whereas level-specific diagnostics allow for trustworthy examination of collinearity at each level.

Simulation Study

We used simulation to further demonstrate our conclusions and to probe additional questions that could not be answered with analytics. Two key goals of the current study are to demonstrate how collinearity's consequences change across different centering specifications, and to clarify how other design factors interact with centering choice to reduce or worsen the consequences of collinearity. Here, we expand on the previous section to further address these goals.

Study Design

We generated multilevel data sets that contained three continuous level-1 predictors, x_{1ij}, x_{2ij}, and x_{3ij}, which followed a multivariate normal distribution, as well as a continuous normally-distributed level-1 outcome, y_{ij}. In all cases, the within-cluster slope of each predictor was 1, and the between-cluster slope of each predictor was –.8. The standardized within-cluster effect ranged from .07 to .11, and the standardized between-cluster effect ranged from –.01 to –.11; standardized effects changed slightly across conditions because the within- and between-cluster variances of the predictors changed as predictor ICCs changed. Additionally, the proportion of total variance in y_{ij} explained by fixed and random effects at both levels (denoted $R_t^{2(fvm)}$ by Rights & Sterba, 2019) ranged from .068 to .378. Explained outcome variance differed across simulation conditions because we opted to instead hold the total variance of y_{ij} constant (see the Avoiding Confounders section below).

In past studies concerning collinearity, it has been consistently found that, all else being equal, the consequences of collinearity are mitigated as sample size increases. Given the highly consistent nature of these findings, we held sample size constant. In all conditions, number of clusters (*J*) was 50, and the size of each cluster (n_j) was randomly sampled from a uniform distribution ranging from 25 to 50. The third predictor (x_{3ij}) was included for comparison, and nothing involving this predictor was manipulated. Data were generated in R, and all models were fit using the *lme4* package (Bates et al., 2015). Code for data generation is available in the Online Code Supplement. Before the main simulation study was run, small calibration studies were conducted to determine the most relevant range of each design factor.

Predictor collinearity was defined as the level-specific pairwise correlations of x_{1ij} and x_{2ij}. Their within-cluster correlation, r_W, and between-cluster correlation, r_B, were varied across the following levels: –.9, –.8, –.7, –.6, 0, .6, .7, .8, and .9. Zero correlations were included to establish a "baseline" condition of no collinearity. Calibration studies showed that .6 was the correlation level at which effects of collinearity became pronounced. From there, we increased correlation by an increment of .10 to maintain granularity. We included both positive and negative correlations due to past work by Mela and Kopalle (2002) which suggests there may be asymmetric effects of collinearity depending on the direction of association. Negative correlations have never been included in past studies concerning collinearity in multilevel data, so whether this asymmetry may carry over to multilevel settings is unknown. Finally, in all conditions, $cor(x_{1ij}, x_{3ij})$ and $cor(x_{2ij}, x_{3ij})$ were held constant at .1 at both levels.

ICC_y was varied across .05 and .3. These represent small and large values that are regularly observed in empirical data. The goal of varying ICC_y was to support our conclusions from the β_{GLS} derivation and provide further clarity given mixed findings in past work. We expected the effects of within-cluster collinearity would be exacerbated by smaller ICC_y and the effects of between-cluster collinearity would be exacerbated by larger ICC_y.

$\text{ICC}_{x_{1ij}}$ and $\text{ICC}_{x_{2ij}}$ were also varied across .05 and .3. Given the importance of $\text{ICC}_{x_{1ij}}$ and $\text{ICC}_{x_{2ij}}$ in Eq. 8, we expected that each would play an interactive role with collinearity to affect conflated point estimates and other aspects of estimation. Predictor ICC has never been varied in past multilevel simulation studies concerning collinearity, so this was another novel aspect of the current study. In all conditions, $\text{ICC}_{x_{3ij}}$ was held constant at .25.

Finally, we varied correlations among the random effects. The correlation of the random intercept and the random slope of x_{1ij} was varied across .1 and .7. Additionally, the correlation of the random slopes of x_{1ij} and x_{2ij} was varied across .1 and .7. In the extant literature concerning collinearity in multilevel models, there is speculation that greater redundancy among the random effects, either independently or interactively with collinearity among observed predictors, may also have consequences for bias and precision (Stinnett, 1994; Zhang & Chen, 2013). No prior studies have manipulated correlations among random effects, so we probed its influence here. Pairwise correlations among all other random effects were held constant at .05.

We implemented a fully crossed design. With 9 levels of r_W, 9 levels of r_B, 2 levels of ICC_y, 2 levels of $\text{ICC}_{x_{1ij}}$, 2 levels of $\text{ICC}_{x_{2ij}}$, 2 levels of correlation across random intercept and slope, and 2 levels of correlation across random slopes, we implemented $(9 \times 9 \times 2 \times 2 \times 2 \times 2 \times 2) =$ 2,592 conditions. For a summary of the simulation conditions, see Table 2. Each condition was repeated until we obtained 1,000 replications in which no convergence warnings or other warnings were produced during model fitting. If a warning was produced, the design conditions and content of the warning were saved but all other output from that replication was discarded.

Models Fit

On each replication, after the multilevel data set was generated, three models were fit to it. The fully conflated model included uncentered x_{1ij}, x_{2ij}, and x_{3ij} as predictors with random slopes (Eq. 18). The fully disaggregated model included cluster mean centered x_{1ij}, x_{2ij}, and x_{3ij} as level-1 predictors with random slopes, and their cluster means were included as level-2 predictors (Eq. 19). In the partially disaggregated model, x_{1ij} was uncentered whereas x_{2ij} and

x_{3ij} were disaggregated (Eq. 20). Each model was fit first with REML estimation, then with full information maximum likelihood (FIML) estimation[2]; these estimation methods are known to yield different random effect (co)variance estimates, so we sought to ascertain whether the effects of predictor collinearity on random effect (co)variance estimates may differ for REML vs. FIML estimation. In all cases, the optimizer was bound optimization by quadratic approximation.

Fully conflated model:

$$\begin{aligned} y_{ij} = {} & \gamma^*_{00} + \gamma^*_{10} x_{1ij} + \gamma^*_{20} x_{2ij} + \gamma^*_{30} x_{3ij} + \\ & u_{1j} x_{1ij} + u_{2j} x_{2ij} + u_{3j} x_{3ij} + u_{0j} + e_{ij} \end{aligned} \tag{18}$$

Fully disaggregated model:

$$\begin{aligned} y_{ij} = {} & \gamma_{00} + \gamma_{10}\left(x_{1ij} - x_{1.j}\right) + \gamma_{20}\left(x_{2ij} - x_{2.j}\right) + \gamma_{30}\left(x_{3ij} - x_{3.j}\right) + \gamma_{01} x_{1.j} + \gamma_{02} x_{2.j} + \gamma_{03} x_{3.j} + \\ & u_{1j}\left(x_{1ij} - x_{1.j}\right) + u_{2j}\left(x_{2ij} - x_{2.j}\right) + u_{3j}\left(x_{3ij} - x_{3.j}\right) + u_{0j} + e_{ij} \end{aligned} \tag{19}$$

Partially disaggregated model:

$$\begin{aligned} y_{ij} = {} & \gamma_{00} + \gamma^*_{10} x_{1ij} + \gamma_{20}\left(x_{2ij} - x_{2.j}\right) + \gamma_{30}\left(x_{3ij} - x_{3.j}\right) + \gamma_{02} x_{2.j} + \gamma_{03} x_{3.j} + \\ & u_{1j} x_{1ij} + u_{2j}\left(x_{2ij} - x_{2.j}\right) + u_{3j}\left(x_{3ij} - x_{3.j}\right) + u_{0j} + e_{ij} \end{aligned} \tag{20}$$

In Eqs. 18-20, γ^* is a conflated fixed effect whereas γ is a level-specific fixed effect. The u terms are random intercepts (u_{0j}) and random slopes (u_{1j}, u_{2j}, u_{3j}) for cluster *j*, whereas e_{ij} is the level-1 residual for observation *i* in cluster *j*. Using the first predictor as an example, x_{1ij} is the uncentered predictor for observation *i* in cluster *j*, $x_{1ij} - x_{1.j}$ is the cluster mean centered predictor, and $x_{1.j}$ is the cluster mean.

Outcome Measures

---

[2] Throughout this paper, full information maximum likelihood (FIML) estimation is synonymous with maximum likelihood estimation (MLE), as the two are equivalent in this case.

On each replication, for each of the three models, we recorded all fixed effect estimates (intercept and slopes), their estimated *SE*s, the estimated covariance matrix of random effects (the $\widehat{\mathbf{T}}$ matrix), and the estimated level-1 residual variance ($\hat{\sigma}^2$).

We used ANOVAs to pinpoint which design factors exacerbated or mitigated the consequences of collinearity in each model. For each outcome of interest, we ran a set of two-way ANOVAs that contained every pairwise crossing between r_W and the other design factors, and between r_B and the other design factors. The pairs of design factors whose ANOVAs yielded the largest effect sizes were then plotted. For each of the three models, outcomes of interest were (1) the fixed effect estimates (which were plotted alongside the level-1 and level-2 data-generating slopes); (2) relative bias in the fixed effect *SE*s, compared to their empirical standard deviations across iterations within each simulation condition; and (3) relative bias in the random effect estimates, compared to the data-generating parameters in each simulation condition.

Avoiding Confounders

In past simulation studies concerning collinearity in multilevel data, the total variance of y_{ij} was not held constant across conditions. When predictor collinearity changes, the explained variance of y_{ij} also changes. If this is not offset, then incidentally, the entire variance of y_{ij} changes. Nonconstant $var(y_{ij})$ has likely been an important confounder in past studies.

In multilevel data, the total variance of y_{ij} can be decomposed into five sources (Rights & Sterba, 2019). (1) *f1* is the level-1 variance of y_{ij} that is explained by level-1 predictors via fixed slopes; (2) v is the level-1 variance of y_{ij} that is explained by level-1 predictors via random slope (co)variation; (3) σ^2 is level-1 residual variance of y_{ij}; (4) *f2* is level-2 variance of y_{ij} that is explained by the fixed slopes of level-2 predictors; (5) τ_{00} is level-2 residual variance of y_{ij} that is attributable to random intercept variation. Both *f1* and *f2* are influenced by the strength of the fixed slopes and the strength of relations among the predictors, at the relevant level. Therefore, changing within-cluster collinearity changes *f1*, whereas changing between-cluster collinearity changes *f2*. Similarly, changing covariation among the random slopes changes *v*. We held the level-1 variance of y_{ij} constant by adjusting σ^2 as needed, commensurate with the changes in *f1* and *v* across simulation conditions. We held the level-2 variance of y_{ij} constant by adjusting τ_{00} as needed according to changes in *f2*. As a result, total $var(y_{ij})$ was held constant across all conditions.

It is likely that spurious results were identified in past work due to nonconstant $var(y_{ij})$ across conditions. However, in this study, we introduced a different confounder in that the proportion of *explained* variance of y_{ij} was different across conditions. To thoroughly examine the impact of this confounder, we ran the simulation study a second time, wherein the explained variance of y_{ij} was held constant at each level and the total variance of y_{ij} was allowed to vary across conditions. We briefly summarize these findings in the Discussion section, and all results from this supplementary study can be found in Online Appendix B. Throughout this paper, we focus on results from the original study wherein the total variance of y_{ij} was held constant.

Results

Fully Conflated Model

Fixed effect estimates were affected by predictor collinearity, along with other design factors. Two-way factorial ANOVAs indicated the following crossings yielded the largest effect sizes: $r_W \times r_B$, $r_W \times$ ICC_y, $r_W \times \text{ICC}_{x_{1ij}}$, $r_W \times \text{ICC}_{x_{2ij}}$, $r_B \times \text{ICC}_y$, $r_B \times \text{ICC}_{x_{1ij}}$, and $r_B \times \text{ICC}_{x_{2ij}}$. Observed patterns supported our conclusions from Eq. 13; however, contrary to expectation, smaller ICC_y exacerbated the effects of both within- and between-cluster collinearity, though its impact was not as pronounced at level 2. In some conditions where predictor ICCs were dissimilar, conflated slope estimates did not fall between the two level-specific data-generating slopes; see Figure 1.

REML and FIML yielded identical fixed effect estimates, which matched those computed with the β_{GLS} formula in all conditions. Given the problematic nature of these fixed effects estimates, which has been demonstrated here and elsewhere (Cronbach & Webb, 1975; Curran & Bauer, 2010; Kenny & La Voie, 1985), we did not probe fixed effect *SE*s or random effect (co)variance estimates. Because most findings are redundant with our conclusions from the β_{GLS} derivation, we did not include all figures here; see Online Appendix A, Figures A3-A6.

Partially Disaggregated Model

We began by examining whether collinearity, together with other design factors, impacted fixed effect estimates. Because x_{1ij} was uncentered and x_{2ij} was disaggregated, we analyzed conflated slope estimates for x_{1ij} and level-specific slope estimates for x_{2ij}. Conflated slope estimates for x_{1ij} were unaffected by predictor collinearity; however, the within- and between-cluster slope estimates for x_{2ij} were affected by predictor collinearity, as well as by other design factors. For the within-cluster slope, ANOVAs indicated that $r_W \times \text{ICC}_y$, $r_W \times \text{ICC}_{x_{1ij}}$, and $r_W \times \text{ICC}_{x_{2ij}}$ should be plotted. For the between-cluster slope of x_{2ij}, we plotted $r_B \times \text{ICC}_{x_{1ij}}$ and $r_B \times \text{ICC}_{x_{2ij}}$.

For the within-cluster slope of x_{2ij}, when ICC_y was small, strong within-cluster collinearity was associated with biased estimates of the within-cluster effect (Figure 2). The direction of this bias corresponded with the direction of r_W. This pattern also emerged when predictor ICCs were large (Figure 3). The true within-cluster effect was 1.0, and mean estimates ranged from .841

(when r_W = –.90, $\mathrm{ICC}_{x_{1ij}}$ = .30, and $\mathrm{ICC}_{x_{2ij}}$ = .30) to 1.127 (when r_W = .90, $\mathrm{ICC}_{x_{1ij}}$ = .30, and $\mathrm{ICC}_{x_{2ij}}$ = .30), which correspond to –15.9% and +12.7% relative bias, respectively.

For the between-cluster slope of x_{2ij}, between-cluster collinearity was associated with biased estimates in all conditions. Large $\mathrm{ICC}_{x_{1ij}}$ and a small $\mathrm{ICC}_{x_{2ij}}$ drastically exacerbated this bias (Figure 4). The true between-cluster effect was –.8, and mean estimates ranged from –10.25 (when r_W = .9, $\mathrm{ICC}_{x_{1ij}}$ = .3, and $\mathrm{ICC}_{x_{2ij}}$ = .05) to 8.91 (when r_W = –.9, $\mathrm{ICC}_{x_{1ij}}$ = .3, $\mathrm{ICC}_{x_{2ij}}$ = .05), which amount to 1,181% and –1,214% relative bias, respectively.

REML and FIML estimates were identical, so Figures 2-4 collapse across results from both estimation methods. We did not examine *SE*s or random effects estimates from this model given the problematic nature of the fixed effect estimates.

Fully Disaggregated Model

Unlike in the conflated and partially disaggregated models, fixed effect estimates in the fully disaggregated model were unaffected by predictor collinearity and were unbiased in all conditions. Therefore, we ran two-way ANOVAs to assess whether predictor collinearity, together with other design factors, impacted relative bias in the *SE*s of the fixed effect estimates. The benchmark used to compute relative bias was the empirical standard deviation of the estimates across the 1,000 iterations of each simulation condition.

Standard Errors. There was a main effect of r_W on relative bias in the *SE*s of the estimated within-cluster effects of x_{1ij} and x_{2ij}. When r_W was further from zero in either direction, relative bias increased. When r_W was –.90, average relative bias (collapsing across all other design factors) was 4.5% for x_{1ij} and 4.4% for x_{2ij}. When r_W was .90, average relative bias was 4.3% for x_{1ij} and 4.6% for x_{2ij}. Focusing on the within-cluster effect of x_{1ij} (as results for x_{1ij} and x_{2ij} were identical), Figure 5 shows empirical standard deviations of the fixed effect estimates and average *SE*s for each value of r_W. Within-cluster collinearity led to a true increase in the sampling variability of level-1 fixed effect estimates (standard deviation increased). However, when r_W was further from zero, that true increase was accompanied by increased relative bias in the *SE*s; in Figure 5, this is indicated by the two lines (one depicting empirical standard deviation and one depicting average estimated *SE*) drifting futher apart as r_W becomes stronger. REML and FIML estimation yielded identical results, so Figure 5 collapses across both.

For relative bias in the *SE*s of estimated between-cluster effects, ANOVAs indicated no main or interactive effects of any design factors. To probe this finding, we plotted empirical standard deviations alongside estimated *SE*s as a function of r_B (Figure 6). Results for REML and FIML differed, so each is shown separately in Figure 6. Between-cluster collinearity resulted in a true increase in the sampling variability of level-2 fixed effect estimates (standard deviation increased). However, increasing r_B was *not* associated with increased relative bias in the estimated *SE*s. Across all levels of r_B, *SE*s showed a downward relative bias of about –7% for REML and –10% for FIML estimation. In Figure 6, the stability of this relative bias is indicated by the lines remaining about the same distance apart across all levels of r_B.

Random Effect (Co)variances. We examined relative bias in the random effect (co)variance estimates as a function of predictor collinearity and other design factors. REML and FIML estimation yielded different results for each, so they are plotted separately in subsequent figures. Beginning with the random intercept variance (τ_{00}), downward relative bias arose in many conditions, and there was a symmetric effect of r_W. When ICC_y was small, relative bias was greater. Across all conditions, downward relative bias was much greater for FIML estimates. When ICC_y was small, relative bias ranged from .7% to 3.2% for REML, and 12.6% to 14.5%

for FIML estimation. When ICC_y was large, REML estimates were unbiased, but relative bias ranged from 7.7% to 7.9% for FIML. See Figure A7 in Online Appendix A.

For bias in the random slope variance estimates for collinear predictors (τ_{11} and τ_{22}), ANOVAs indicated that r_W should be plotted with the ICC of the corresponding predictor (x_{1ij} for τ_{11} and x_{2ij} for τ_{22}). See Figure 7. When r_W was further from zero in either direction, there was greater upward relative bias. Large predictor ICC exacerbated this bias. REML resulted in greater bias than FIML, and this difference became more pronounced as r_W was further from zero. For both estimation methods and both variances, the greatest relative bias was observed when r_W was –.90 and the relevant predictor ICC was large (.30). With REML, relative bias reached a maximum of 352% for τ_{11} and 342% for τ_{22}. With FIML, relative bias in τ_{11} reached a maximum of 317% and 307% for τ_{22}.

Random slope variance estimates for the non-collinear predictor (τ_{33}) showed upward relative bias in most conditions. There was an asymmetric effect of r_W such that negative r_W led to greater bias than positive r_W. When ICC_y was small, bias was exacerbated. In all conditions, REML estimates were more biased. Relative bias was greatest when r_W was –.90 and ICC_y was .05, and reached 41.9% for REML and 31.6% for FIML. See Figure A8 in Online Appendix A.

Random slope covariance estimates for the collinear predictors (τ_{21}) were upwardly biased when r_W was negative, and downwardly biased when r_W was positive. Bias was exacerbated by small ICC_y and large ICC of either predictor. There was also an interaction with the correlation of the random slopes of x_{1ij} and x_{2ij}; when this correlation was smaller, bias was exacerbated. REML and FIML estimates were similar when collinearity was low-to-moderate, but at the strongest levels of collinearity, REML yielded greater relative bias than FIML. In the conditions that yielded the greatest relative bias (small ICC_y, large predictor ICCs, and small random slope correlation), r_W = –.90 was associated with relative biases of 3,986% for REML and 3,651% for FIML, whereas r_W = .90 was associated with relative biases of –3,132% for REML and –2,985% for FIML. See Figures A9-A11 in Online Appendix A.

We also probed bias in estimated random slope covariances where one predictor was involved in the collinearity and the other was not (τ_{31} and τ_{32},), as well as covariances among the random intercept and random slopes (τ_{10}, τ_{20}, τ_{30}). In general, relative bias was larger when r_W was further from zero, ICC_y was small, and predictor ICCs were large. See Figures A12-A15.

Nonconvergence. Findings were extremely similar for REML and FIML, so the following results collapse across the two estimation methods. In total, .4% of iterations yielded a nonconvergence warning. Strong within-cluster collinearity was associated with greater incidence of nonconvergence; however, nonconvergence remained stable across all levels of between-cluster collinearity (Online Appendix A, Figure A16). Larger ICC_y was associated with more nonconvergence warnings; 40.11% of warnings arose when ICC_y was .05, whereas 59.9% of warnings arose when ICC_y was .30. Nonconvergence rate did not differ across levels of predictor ICC or covariation the random effects. For more detail about rates of nonconvergence across simulation conditions, see Online Appendix A Figures A17-A19.

Discussion

The purpose of this study was to integrate the topics of centering and collinearity in multilevel models. To this end, we reconciled the broad literatures on each of these topics, reviewed level-specific and conflated correlation and covariance, and formalized relationships between predictor collinearity and multilevel model parameter estimates. We demonstrated that predictor collinearity, both within and between clusters, deterministically impacts point estimates in the fully conflated multilevel model and in the partially disaggregated multilevel model. These results are a key departure from single-level regression, where collinearity affects estimates' precision, not bias. In multilevel models, in the presence of contextual effects (i.e., when the within- and between-cluster effect of a predictor are not equal) we have proven that point estimates themselves can be biased by predictor collinearity if full disaggregation is not conducted.

Of note, our analytic conclusions are generalizable to both continuous and categorical predictors. Use of a multicategorical predictor necessitates the inclusion of multiple coding variables (e.g., dummy, contrast, or effect codes). Coding variables will necessarily be correlated, and their level-specific correlations may differ greatly depending on group proportions within and across clusters. The current work suggests that slope estimates associated with coding variables may easily become biased if full disaggregation is not conducted, reiterating its importance in the presence of categorical predictors.

With respect to the fully conflated model, our conclusions from the maximally general β_{GLS} derivation were mostly supported by the simulation study. Conflated slope estimates varied as a function of predictor collinearity at each level. As expected, smaller ICC_y exacerbated the effect of within-cluster collinearity on the point estimates. Contrary to expectation, smaller ICC_y also exacerbated the effect of between-cluster collinearity; this finding is likely attributable to the more complex weight, which is not directly informed by ICC_y, associated with between-cluster collinearity in the β_{GLS} formula.

Predictor ICCs interacted with collinearity in interesting ways to "push around" conflated slope estimates. When within-cluster collinearity was strong and predictor ICCs were dissimilar, conflated slope estimates did not lie between the two level-specific slopes. Notably, this finding coincides with Eq. 8, which shows that dissimilar predictor ICCs is a necessary data condition for a conflated correlation (r_T) that does not lie between r_W and r_B (also see Online Appendix A). Overall, dissimilar predictor ICCs can yield both conflated slopes and r_Ts that do not lie between the two level-specific quantities by which they are informed. Therefore, dissimilar predictor ICCs are associated with particularly misleading results when predictors are not disaggregated. Past work has not identified data conditions wherein it is possible to obtain a conflated slope estimate that does not lie between the two level-specific slopes. In fact, it was understood that a conflated slope would always lie between the two, according to Raudenbush and Willms' (1995) expression of the conflated slope as a weighted average of within- and between-cluster effects. It appears that this pattern need not hold in the context of multiple predictors that are collinear.

To wrap up our discussion of the fully conflated model, we acknowledge that in some situations, researchers may conclude that the conflated model is appropriate if the within- and between-cluster effects of a predictor are found to be (nearly) equal. However, even in these cases, the conflated slope estimate is still systematically "pushed around" by a variety of data characteristics, and is still uninterpretable. Slopes are interpreted in terms of what happens in response to a "one-unit change" in a predictor, and if "one-unit change" means something different within and across clusters (it must), then equal slopes are only incidentally equal, and their equality is substantively meaningless.

Findings concerning the partially disaggregated model were also novel. Such a model has never been investigated in studies concerning multilevel collinearity, so these results are important to unpack. When two predictors are collinear, and one is uncentered whereas the other is disaggregated, fixed effect estimates with respect to the *disaggregated* predictor will be biased – sometimes drastically so – as a result of collinearity. This finding has key practical implications. Many applied users of MLM understand the importance of disaggregation for the interpretability of estimates; however, predictors whose effects are not of key substantative interest, such as covariates, are frequently left uncentered. Here we show that when predictors are collinear and some are left uncentered whereas other, likely the most substantively important, predictors are disaggregated, that disaggregation will *not* yield unbiased estimates. If predictors are collinear to even a moderate degree (e.g., $r = .6$), then slope estimates associated with the disaggregated predictor will be untrustworthy. It is therefore of utmost importance to disaggregate all predictors in the model, even those that are not substantively important.

In the fully disaggregated model, slope estimates were unaffected by collinearity. Collinearity's consequences for this model most closely mimic those observed in single-level regression, such that fixed effect point estimates were unaffected but their *SE*s were enlarged. However, enlarged *SE*s were accompanied by upward bias. This is an important departure from single-level regression, where large *SE*s are not the result of upward bias, but are an accurate reflection of the increased sampling variability of each point estimate. In the fully disaggregated multilevel model, collinearity led to *both* true increases in the sampling variability of the point estimates *and* upward bias in the *SE*s at level 1. This suggests that for *SE*s, the consequences of collinearity are more severe in multilevel models than in single-level models. This finding also clarifies confusion from past multilevel studies, as some reported upwardly biased *SE*s whereas others did not. The upward bias observed here could be classified as negligible according to some criteria (e.g., <5%); perhaps some studies observed similar patterns but reported a lack of bias due to its small magnitude. Additionally, prior work is often unclear as to which centering specification was under study. Relative bias in the *SE*s varied as a function of collinearity for within-cluster estimates, but not for between-cluster estimates, so it is possible that this result was muddied in past work that did not conduct disaggregation.

This study provides compelling evidence regarding the consequences of predictor collinearity for random effect (co)variance estimates, which were previously poorly understood due to scarce research. Within-cluster collinearity led to bias in nearly every random effect (co)variance estimate, whereas between-cluster collinearity did not affect bias. Small ICC_y and large predictor ICCs exacerbated bias. Across negative vs. positive predictor correlations, bias tended to be symmetric for variances but asymmetric for covariances. Collinearity caused the greatest relative bias in the random slope (co)variance estimates corresponding to the collinear predictors. Relative bias tended to increase exponentially across the collinearity levels that we tested, suggesting that at low-to-moderate predictor collinearity, negligible bias could be maintained.

Our results concerning (co)variance estimates differ notably from past findings, which are sparse. Shieh and Fouladi (2003) reported small-to-moderate downward bias in all random effect (co)variances except random intercept variance, whereas Yu et al. (2015) reported no bias in any data conditions. We observed upward bias in all random slope variance estimates, and both upward and downward bias in all covariance estimates, which was sometimes extreme. Our divergent findings may be attributable to important, yet previously unaddressed, differences in data generation procedures and model specification. Past studies were often unclear about the centering method used, and the total variance of y_{ij} was not held constant. Therefore, it is unclear whether the same random effects are being compared; random effects in the conflated model suffer from similar problems as the fixed effects (Rights, 2022), so examination of their bias may be misleading. It is also likely that nonconstant variance of y_{ij} across conditions confounded prior findings. In future work, transparency regarding data-generation processes and model specification details will be very important for facilitating replicable results and an accumulation of evidence regarding collinearity's consequences for multilevel models.

In the fully disaggregated model, differences emerged between REML and FIML estimation with respect to *SE*s and random effect (co)variance estimates. *SE*s of the level-2 fixed effect estimates were less biased under REML; this finding was expected (Raudenbush & Bryk, 2002) and was consistent across all collinearity conditions. Random effect (co)variance estimates also differed, especially at the most extreme levels of collinearity. In these conditions, REML tended to yield slightly greater bias than FIML, including for all random slope (co)variances. This was surprising given that REML typically yields less biased variance estimates than FIML, whose estimates tend to be downwardly biased. (We observed this pattern only for $\hat{\tau}_{00}$.) Differences between REML and FIML were slight for most (co)variance estimates when collinearity was small-to-moderate. However, it was unexpected that REML was often more sensitive to severe predictor collinearity. Future work should explore whether this pattern holds for smaller samples. Overall, because neither estimation method was consistently superior over the other, collinearity conditions in the data – together with researcher priorities regarding which estimates are most important to obtain with the least bias – should inform the choice between REML and FIML.

There was an asymmetric effect of predictor collinearity on many outcomes. In past studies of collinearity in multilevel models, only positive predictor correlations have been included, so these findings are novel. In all models, it appeared that many outcomes were more strongly impacted by negative r_W than by positive r_W. Interestingly, this pattern was *not* observed as consistently in the second simulation study, when *explained* $var(y_{ij})$ was held constant rather than *total* $var(y_{ij})$. See Online Appendix B. In the original simulation study, explained $var(y_{ij})$ varied across conditions, which likely contributed to asymmetric findings. In our simulated data sets, x_{1ij} and x_{2ij} each exerted a positive effect on y_{ij} at level 1. Therefore, as r_W became more positive, more $var(y_{ij})$ was explained at level 1, and as r_W became more negative, less $var(y_{ij})$ was explained at level 1. (Had x_{1ij} and x_{2ij} each exerted a negative effect on y_{ij}, this pattern would have reversed.) It is known that the harmful effects of collinearity tend to be exacerbated when the explained variance of the outcome is smaller (e.g., Hayo, 2018; O'Brien, 2007). Thus, it makes sense that negative r_W tended to have a "stronger" effect on many outcomes, because less $var(y_{ij})$ was explained in these conditions. In summary, the effects of negative vs. positive predictor correlation will differ according to the direction of their relations with the outcome, and there is not sufficient evidence to suggest that negative vs. positive predictor correlation will universally have greater or lesser effects on estimation.

Another key goal of this study was to demonstrate the benefits of disaggregation for the diagnosis of collinearity in multilevel data. The utility of level-specific collinearity diagnostics was demonstrated in many ways. First, diagnostics at each level are independent of each other, allowing for trustworthy examination of collinearity at each level in isolation. Second, conflated collinearity diagnostics are misleading and arbitrary, highlighting their lack of utility. Third, level-specific collinearity diagnostics identify the collinearity that informs parameter estimates; level-specific collinearity impacts bias and precision in all models, so examining collinearity among conflated predictors is not useful for understanding how estimation has been impacted. Both mathematically and conceptually, level-specific diagnostics possess major advantages and are necessary for accurate diagnosis of collinearity in multilevel data.

The benefits of disaggregation follow from the fact that it yields two sets of level-specific predictors that are mutually independent. Collinearity can arise across levels if a level-1 predictor is left uncentered, as the uncentered level-1 predictor may be correlated with its level-2 cluster mean or with other level-2 predictors; however, this collinearity can be eliminated by disaggregation. This property exactly mimics *non-essential collinearity* in single-level regression, where non-essential collinearity is due solely to the scaling of variables and can be removed through centering (Cohen et al., 2003; Dalal & Zicker, 2012). Thus, cross-level collinearity should be considered *non-essential* in multilevel settings.

Limitations of the study and opportunities for future work are important to address. First, to keep the number of simulation conditions manageable, we did not vary all relevant design factors. Sample size at each level was held constant, and our simulated data sets had many clusters and large cluster sizes, so we likely obtained conservative estimates of the harmful effects of collinearity. Additionally, the differences between FIML and REML may have been more pronounced had sample sizes been smaller. In the future, it may be useful to replicate our findings on smaller samples and explore how the consequences of collinearity, and interactions with other design factors and estimation methods, may be more pronounced. We also held slopes constant at each level. Strength of the relations between predictors and the outcome influences the consequences of collinearity, such that stronger relations (i.e., a higher model R^2) mitigate collinearity problems (Hayo, 2018; O'Brien, 2007). This finding likely carries over to multilevel settings (Clark, 2013), but there has not been an accumulation of evidence. Second, we included diagnostics primarily for demonstrative purposes (i.e., to show they are misleading when applied to uncentered predictors). Deeper study of diagnostics' behavior on multilevel data, and how they may be adapted to the unique challenges that arise in multilevel models is an important topic for future study. For example, in what is often termed the "contextual effects" multilevel model, level-1 predictors are uncentered and their cluster means are included as level-2 predictors; this model yields within-cluster effects at level 1 alongside "contextual effects" at level 2, yet the level-1 and level-2 predictors are not orthogonal. It remains unclear, but will be useful to explore, how collinearity diagnostics should be adapted to such a situation. Third, future work may explore how to mitigate collinearity when higher-order terms are included in multilevel models. Cluster mean centered predictors that are raised to a power, or involved in interactions, incidentally contain some level-2 variance (Rights & Sterba, 2021) and further steps may need to be taken to ensure collinearity problems do not arise.

Finally, although our focus was the identification of collinearity in multilevel data and understanding its implications for multilevel models, it is also important to discuss approaches for remediating collinearity problems once they have been identified. Many recommendations have been posed in single-level regression; for example, a set of predictors may be replaced with a smaller set of principal component scores, predictors that exhibit pairwise correlations greater than some threshold may be removed, or factor analysis may be undertaken (e.g., Beckstead, 2012; Dormann et al., 2013). However, many of these remedies become exceedingly more complicated in multilevel contexts, and very little research has probed their performance in multilevel settings. Existing recommendations are to collect more data, redefine the research problem, or respecify the model (Yu et al., 2015), which in most applications would not be feasible or would compromise the theoretical justification of the model. Removing the most egregiously collinear predictor(s) may be the simplest approach in multilevel settings, and while this is often not ideal, disaggregation lends some important benefits if this approach must be taken. If a predictor has been disaggregated, we can identify whether it is problematically collinear at one or both levels. Because level-specific components of predictors are mutually independent, they both need not be included in the model; we can remove one or the other component without compromising estimates or their interpretability (Raudenbush, 2009). If collinearity is problematic, then with disaggregation we can discard less information.

This study is the first to thoroughly integrate the topics of centering and collinearity in multilevel data. By doing so, we developed novel insights into the consequences of collinearity in multilevel models across different centering specifications, as well as the importance of disaggregation for collinearity diagnosis in multilevel data. Evidence points to the necessity of disaggregation as a means to mitigate the consequences of collinearity. However, even if full disaggregation is conducted, we identified some data conditions where strong collinearity may still be harmful for *SE*s and random effect (co)variance estimates. In all, we hope to equip readers with the tools and knowledge to meaningfully identify collinearity in multilevel data, and to keep collinearity problems to a minimum through appropriate model specification.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Beckstead, J. W. (2012). Isolating and examining sources of suppression and multicollinearity in multiple linear regression. *Multivariate Behavioral Research*, *47*(2), 224-246.

Belsley, D. A. (1991). A guide to using the collinearity diagnostics. *Computer Science in Economics and Management*, *4*(1), 33-50.

Bickel, R. (2007). Grand-mean centering, group-mean centering, and raw scores compared. In *Multilevel analysis for applied research: It's just regression!* Guilford Press.

Blaze, T. J., & Ye, F. (2012). The effect of within-and cross-level multicollinearity on parameter estimates and standard errors in multilevel modeling with different centering methods.

In *American Educational Research Association Conference, Vancouver, British Columbia, Canada*.

Bonate, P. L. (1999). The effect of collinearity on parameter estimates in nonlinear mixed effect models. *Pharmaceutical Research*, *16*(5), 709-717.

Clark, P. C. (2013). The effects of multicollinearity in multilevel models (Doctoral dissertation). Wright State University.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

Cronbach, L. J., & Webb, N. (1975). *Between-class and within-class effects in a reported aptitude* treatment interaction: Reanalysis of a study by GL Anderson.*

Curran, P. J., & Bauer, D. J. (2010). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*(1), 583–619.

Dalal, D. K., & Zickar, M. J. (2012). Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organizational Research Methods*, *15*(3), 339-362.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27-46.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods*, *12*(2), 121-138.

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: The problem revisited. *Review of Economics and Statistics*, 49, 92-107.

Gale, L. C. (1987). *Correlations for multilevel data* (Doctoral dissertation). Stanford University.

Goldberger, A. S. (1991). *A course in econometrics.* Harvard University Press.

Gunst, R. F., & Mason, R. L. (1977). Advantages of examining multicollinearities in regression analysis. *Biometrics, 33*(1), 249-260.

Hayo, B. (2018). On standard-error-decreasing complementarity: why collinearity is not the whole story. *Journal of Quantitative Economics*, *16*(1), 289-307.

Hendrickx, J. (2018). *Collinearity in mixed models*. Paper presented at PHUSE EU Connect Conference, Frankfurt, Germany.

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management*, *24*(5), 623–641.

Kenny, D. A., & La Voie, L. (1985). Separating individual and group effects. *Journal of Personality and Social Psychology*, *48*(2), 339.

Kreft, I. G., & De Leeuw, J. (1998). *Introducing multilevel modeling*. Sage.

Kubitschek, W. N., & Hallinan, M. T. (1999). Collinearity, bias, and effect size: Modeling "the" effect of track on achievement. *Social Science Research*, *28*(4), 380-402.

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*(3), 203-229.

Mason, C. H., & Perreault Jr, W. D. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research*, *28*(3), 268-280.

Mela, C. F., & Kopalle, P. K. (2002). The impact of collinearity on regression analysis: the asymmetric effect of negative and positive correlations. *Applied Economics*, *34*(6), 667-677.

Muthén, B. O. (1990). *Mean and covariance structure analysis of hierarchical data.* Los Angeles, CA: UCLA Statistics Series No. 62.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*(5), 673-690.

Raudenbush, S. W. (2009). Adaptive centering with random effects: An alternative to the fixed effects model for studying time-varying treatments in school settings. *Education Finance and Policy*, *4*(4), 468–491.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.

Raudenbush, S. W., & Willms, J. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, *20*(4), 307-335.

Rights, J. D. (2022). Aberrant distortion of variance components in multilevel models under conflation of level-specific effects. *Psychological Methods.* Advance online publication.

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*, *24*(3), 309-338.

Rights, J. D., & Sterba, S. K. (2021). Effect size measures for longitudinal growth analyses: Extending a framework of multilevel model R-squareds to accommodate heteroscedasticity, autocorrelation, nonlinearity, and alternative centering strategies. *New Directions for Child and Adolescent Development, 175,* 65-110.

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

Scott, A. J., & Holt, D. (1982). The effect of two-stage sampling on ordinary least squares methods. *Journal of the American Statistical Association*, *77*(380), 848-854.

Shieh, Y. Y., & Fouladi, R. T. (2003). The effect of multicollinearity on multilevel modeling parameter estimates and standard errors. *Educational and Psychological Measurement*, *63*(6), 951-985.

Silvey, S. D. (1969). Multicollinearity and imprecise estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, *31*(3), 539-552.

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

Stinnett, S. S. (1994). *Collinearity in mixed models*. (Doctoral dissertation). The University of North Carolina at Chapel Hill.

Willan, A. R., & Watts, D. G. (1978). Meaningful multicollinearity measures. *Technometrics*, *20*(4), 407-412.

Yaremych, H. E., Preacher, K. J., & Hedeker, D. (2021). Centering categorical predictors in multilevel models: Best practices and interpretation. *Psychological Methods*. Advance online publication.

Yu, H., Jiang, S., & Land, K. C. (2015). Multicollinearity in hierarchical linear models. *Social Science Research*, 53, 118-136.

Zhang, G., & Chen, J. J. (2013). Adaptive fitting of linear mixed-effects models with correlated random effects. *Journal of Statistical Computation and Simulation*, *83*(12), 2291-2314.

Table 1

Correlation matrices and collinearity diagnostics for uncentered and level-disaggregated predictors across three data-generating conditions

Predictor set	Correlation matrix							VIFs	Condition numbers ($\boldsymbol{\kappa s}$)
Condition 1 (r_W = 0, r_B = 0, ICC$_{x_{1ij}}$= ICC$_{x_{2ij}}$= ICC$_{x_{3ij}}$= .25)									
		x_{1ij}	x_{2ij}	x_{3ij}				x_{1ij}: 1.002	1.053
Uncentered	x_{1ij}	1						x_{2ij}: 1.002	
	x_{2ij}	**-.031**	1					x_{2ij}: 1.002	
	x_{3ij}	.036	.030	1					
		x_{1i}	x_{2i}	x_{3i}	$x_{1.j}$	$x_{2.j}$	$x_{3.j}$	x_{1i}: 1.002	All: 1.158
	x_{1i}	1						x_{2i}: 1.001	Level 1: 1.051
	x_{2i}	**-.023**	1					x_{3i}: 1.002	Level 2: 1.158
Disaggregated	x_{3i}	.041	.021	1				$x_{1.j}$: 1.009	
	$x_{1.j}$	0	0	0	1			$x_{2.j}$: 1.022	
	$x_{2.j}$	0	0	0	**-.095**	1		$x_{3.j}$: 1.013	
	$x_{3.j}$	0	0	0	-.016	.112	1		
Condition 2 (r_W = .7, r_B = -.9, ICC$_{x_{1ij}}$= ICC$_{x_{2ij}}$= ICC$_{x_{3ij}}$= .25)									
		x_{1ij}	x_{2ij}	x_{3ij}				x_{1ij}: 1.353	1.763
Uncentered	x_{1ij}	1						x_{2ij}: 1.352	
	x_{2ij}	**.509**	1					x_{2ij}: 1.005	
	x_{3ij}	.066	.060	1					
Disaggregated		x_{1i}	x_{2i}	x_{3i}	$x_{1.j}$	$x_{2.j}$	$x_{3.j}$	x_{1i}: 2.080	All: 4.687
	x_{1i}	1						x_{2i}: 2.081	Level 1: 2.494

	x_{2i}	**.720**	1					x_{3i}: 1.008	Level 2: 4.687
	x_{3i}	.083	.085	1				$x_{1.j}$: 5.959	
	$x_{1.j}$	0	0	0	1			$x_{2.j}$: 5.995	
	$x_{2.j}$	0	0	0	**-.905**	1		$x_{3.j}$: 1.085	
	$x_{3.j}$	0	0	0	-.036	-.086	1		
Condition 3 (r_W = .25, r_B = .95, ICC$_{x_{1ij}}$= .8, ICC$_{x_{2ij}}$= .01, ICC$_{x_{3ij}}$= .25)									
		x_{1ij}	x_{2ij}	x_{3ij}				x_{1ij}: 1.019	1.153
	x_{1ij}	1						x_{2ij}: 1.007	
Uncentered	x_{2ij}	**.073**	1					x_{2ij}: 1.016	
	x_{3ij}	.120	.046	1					
		x_{1i}	x_{2i}	x_{3i}	$x_{1.j}$	$x_{2.j}$	$x_{3.j}$	x_{1i}: 1.073	All: 6.864
	x_{1i}	1						x_{2i}: 1.070	Level 1: 1.311
	x_{2i}	**.254**	1					x_{3i}: 1.006	Level 2: 6.864
Disaggregated	x_{3i}	.074	.048	1				$x_{1.j}$: 11.328	
	$x_{1.j}$	0	0	0	1			$x_{2.j}$: 11.380	
	$x_{2.j}$	0	0	0	**.955**	1		$x_{3.j}$: 1.111	
	$x_{3.j}$	0	0	0	.309	.316	1		

Note. x_{1ij} = uncentered predictor. x_{1i} = level-1 component of disaggregated predictor. $x_{1.j}$ = level-2 component of disaggregated predictor. See Online Code Appendix for example R code to compute diagnostics.

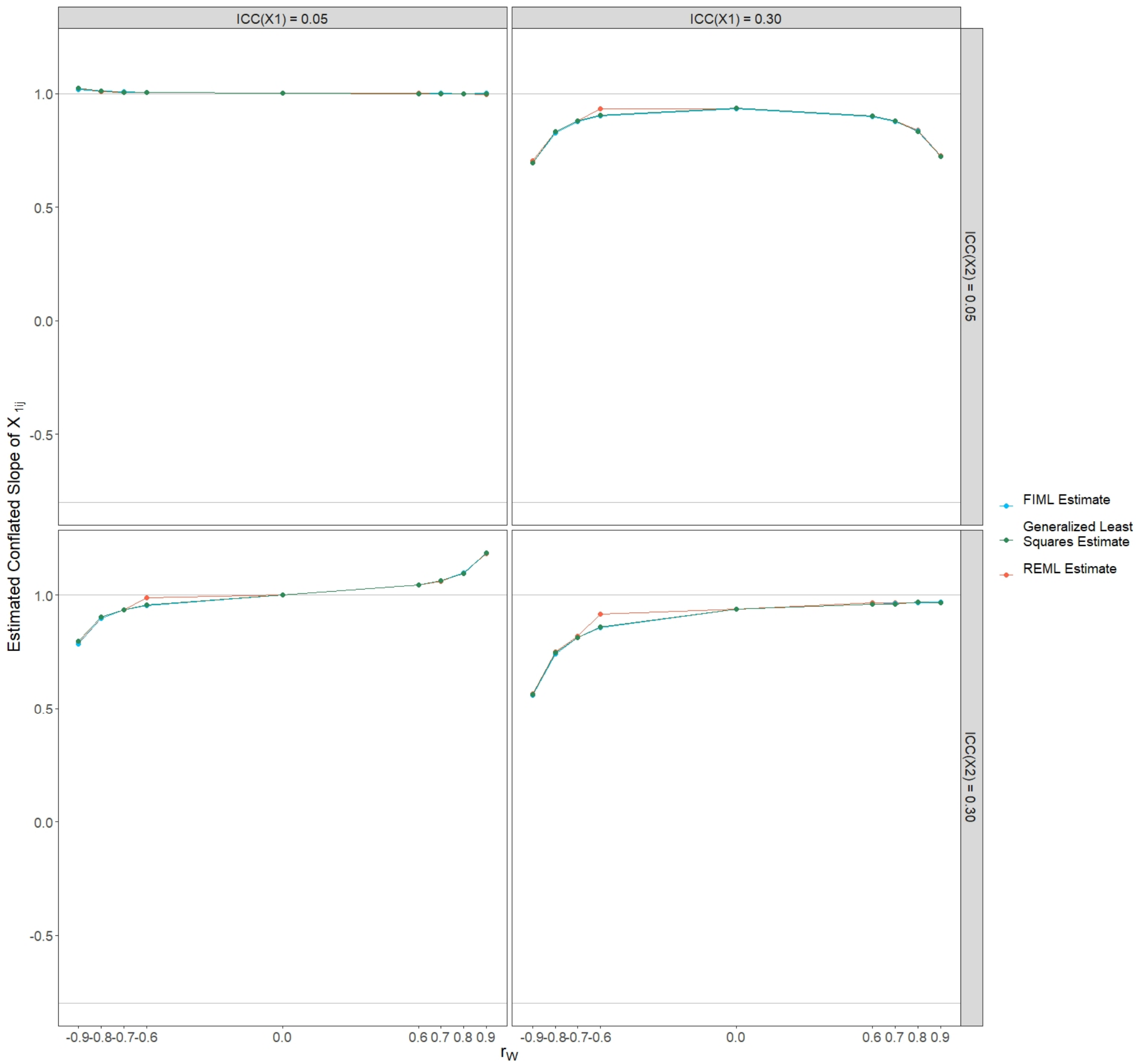
Table 2

Summary of simulation study conditions

Design Factor	**Levels**
Within-cluster correlation of x_{1ij} and x_{2ij} (r_W)	–.9, –.8, –.7, –.6, 0, .6, .7, .8, .9
Between-cluster correlation of x_{1ij} and x_{2ij} (r_B)	–.9, –.8, –.7, –.6, 0, .6, .7, .8, .9
ICC_y	.05, .3
$\text{ICC}_{x_{1ij}}$	.05, .3
$\text{ICC}_{x_{2ij}}$	.05, .3
Correlation of random intercept and random slope of x_{1ij}	.1, .7
Correlation of random slopes of x_{1ij} and x_{2ij}	.1, .7

Figure 1

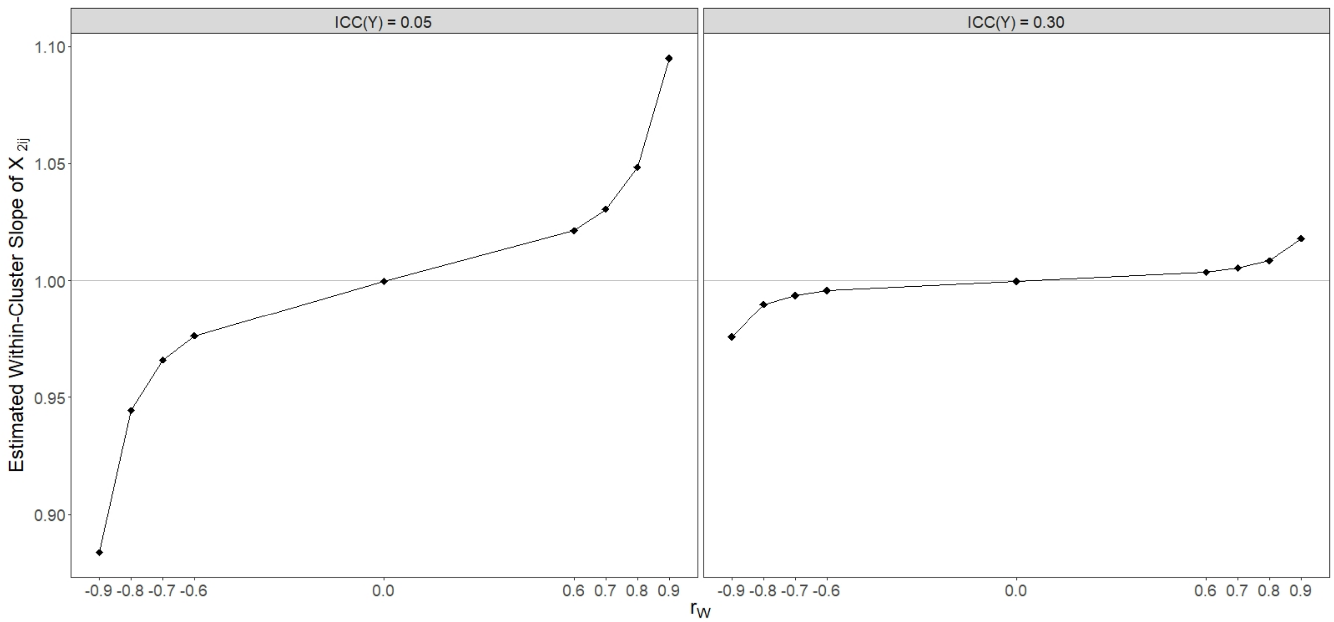
Mean estimates of the conflated slope of x_{1ij} in the fully conflated model, as a function of r_W and predictor ICCs



Note. Solid grey lines at 1.00 and –0.80 are the data-generating within-cluster and between-cluster slope, respectively, of x_{1ij}.

Figure 2

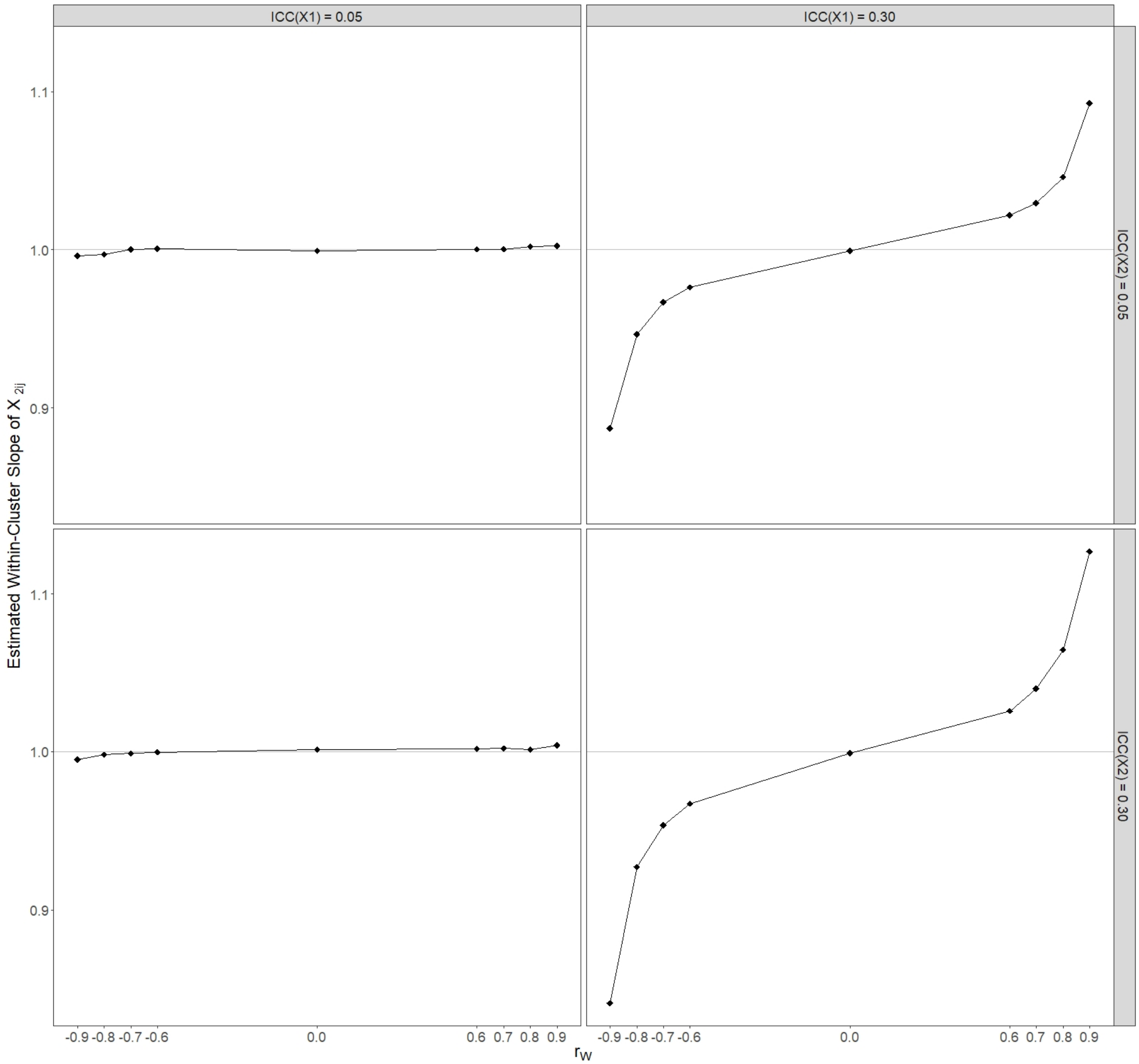
Mean estimates of the within-cluster slope of x_{2ij} in the partially disaggregated model, as a function of r_W and ICC_y



Note. The solid grey line at 1.00 is the data-generating within-cluster slope of x_{2ij}.

Figure 3

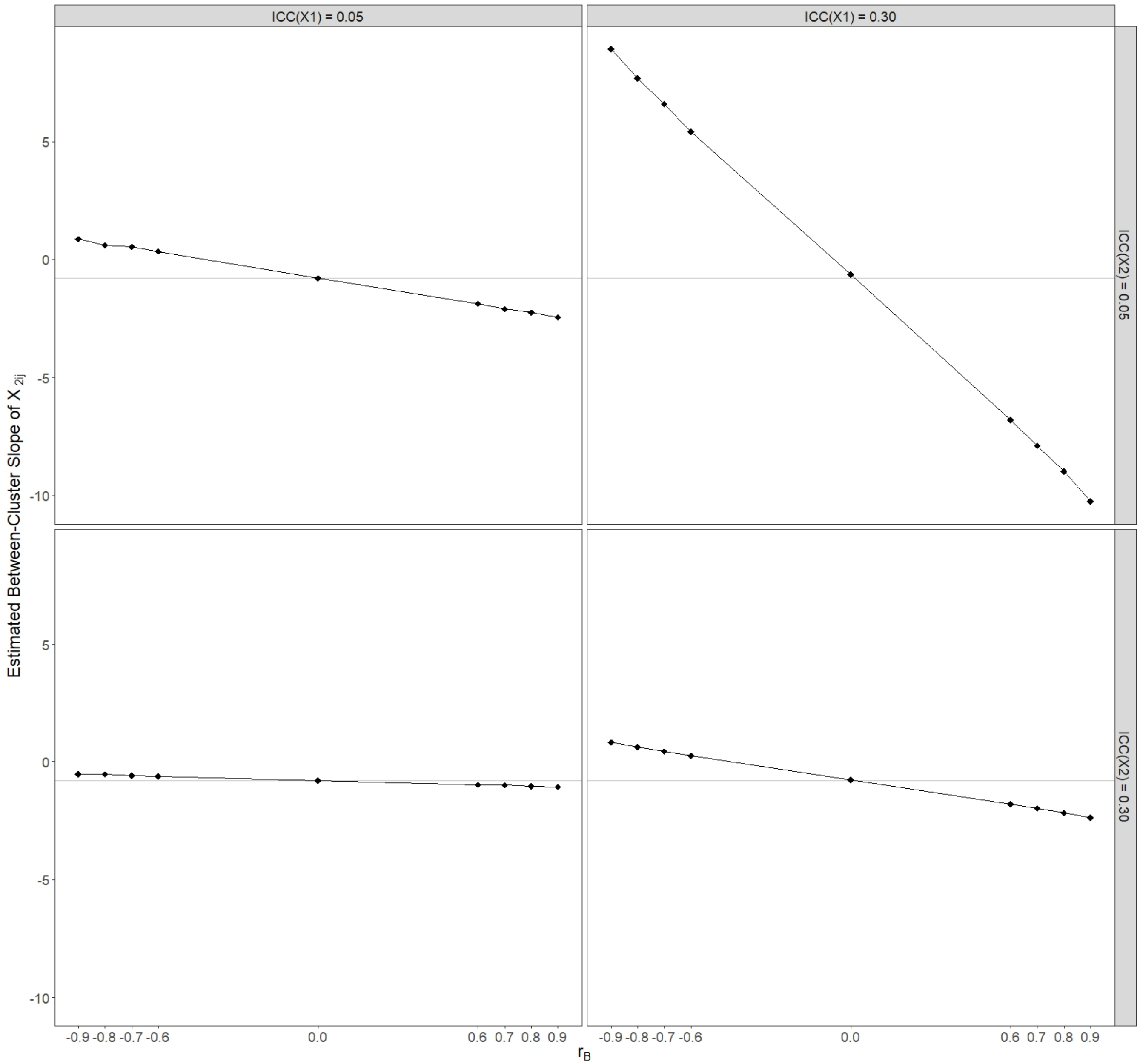
Mean estimates of the within-cluster slope of x_{2ij} in the partially disaggregated model, as a function of r_W and predictor ICCs



Note. The solid grey line at 1.00 is the data-generating within-cluster slope of x_{2ij}.

Figure 4

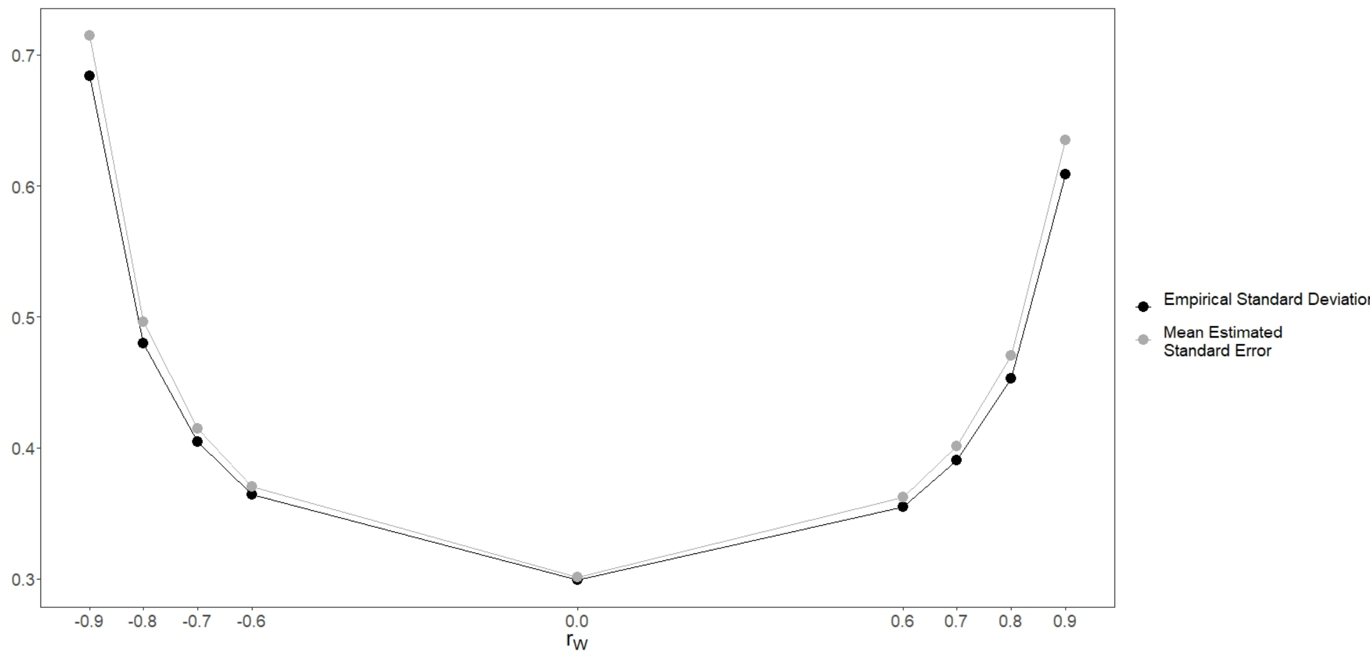
Mean estimates of the between-cluster slope of x_{2ij} in the partially disaggregated model, as a function of r_B and predictor ICCs



Note. The solid grey line at –0.80 is the data-generating between-cluster slope of x_{2ij}.

Figure 5

Empirical standard deviations and estimated standard errors of the within-cluster effect of x_{1ij} *in the fully disaggregated model*



Note. This figure collapses across REML and FIML estimation, as they yielded identical results.

Figure 6

Empirical standard deviations and estimated standard errors of the between-cluster effect of x_{1ij} *in the fully disaggregated model*

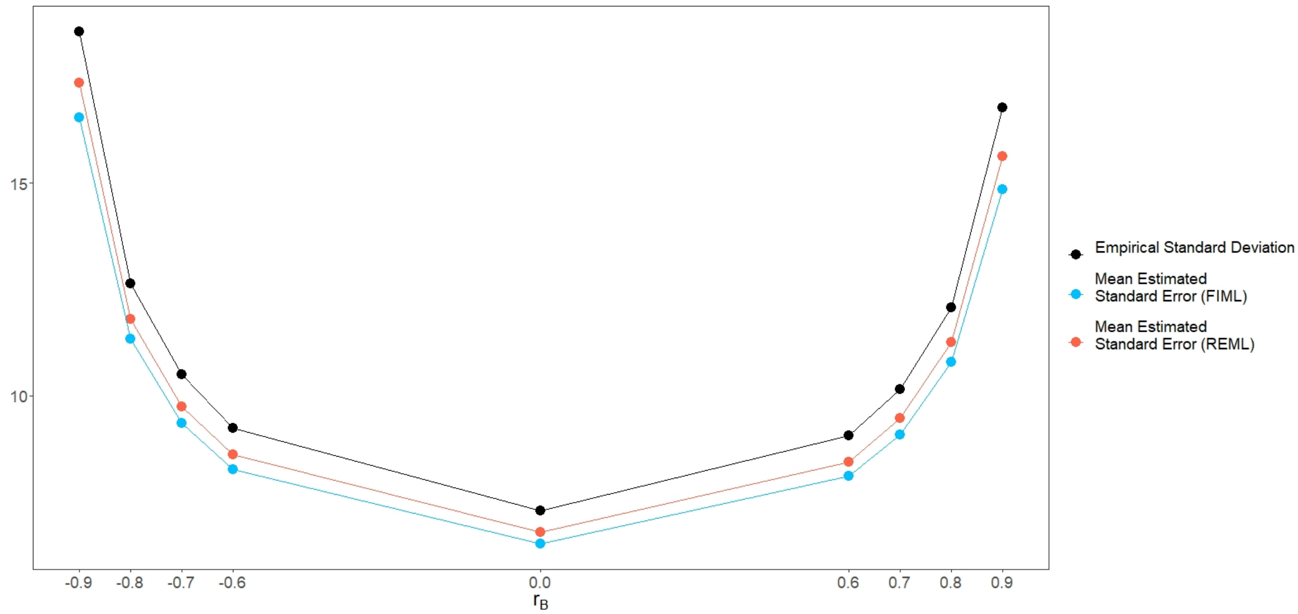
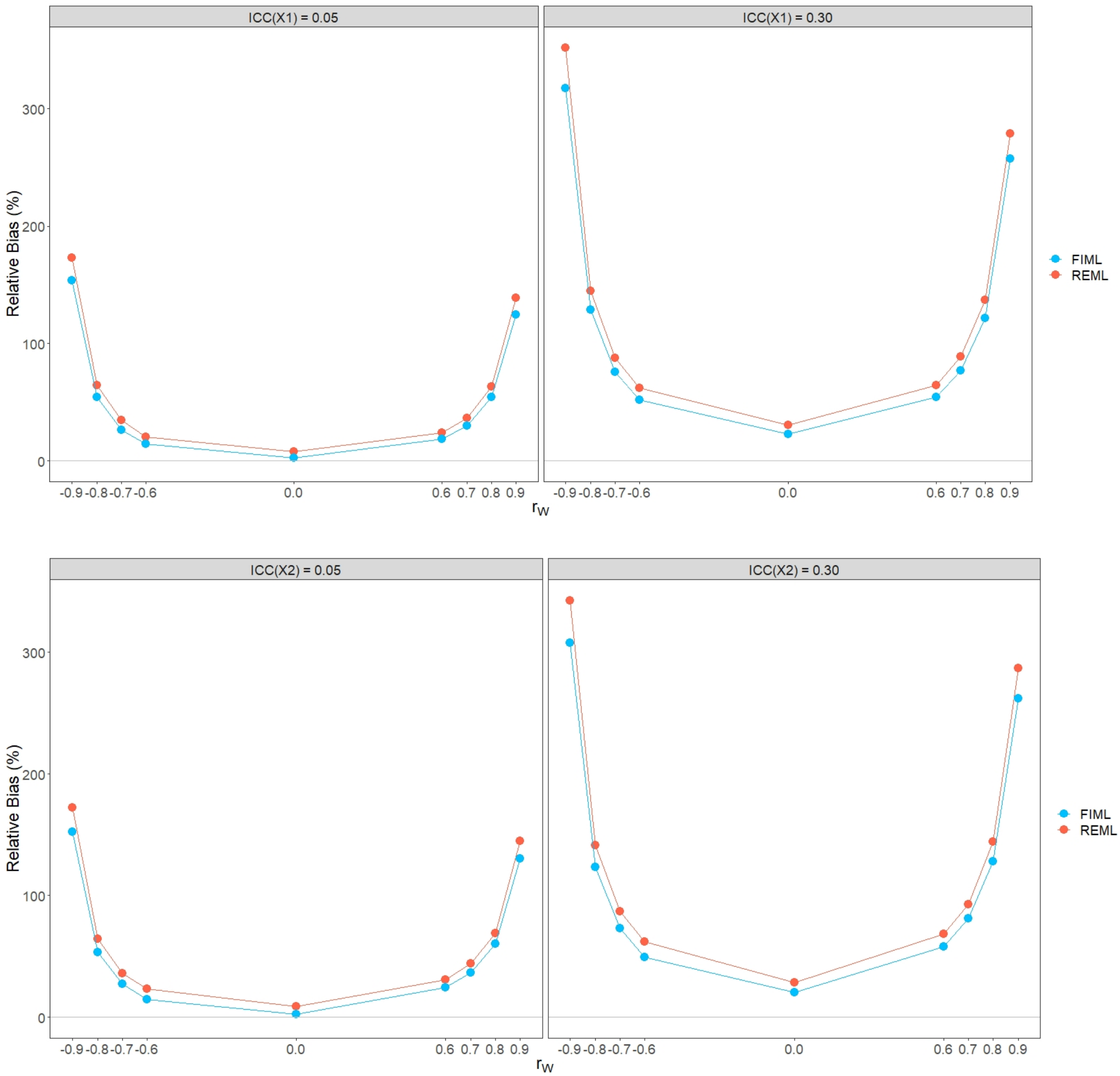


Figure 7

From the fully disaggregated model, relative bias in the estimates of τ_{11} as a function of r_W and $ICC_{x_{1ij}}$ (top panel) and relative bias in the estimates of τ_{22} as a function of r_W and $ICC_{x_{2ij}}$ (bottom panel)



Appendix A: Demonstrating correspondence between the total correlation coefficient and Muthén's standardized total covariance matrix

Gale (1987) and Snijders and Bosker (2012) present an equation for the total correlation coefficient, r_T , from a within-cluster correlation, r_W, and a between-cluster correlation, r_B . This formula pertains to only one correlation at a time and has not been scaled up to a full correlation matrix, $\mathbf{R_T}$.

Additionally, Muthén (1990) showed that the total covariance matrix for multilevel data, $\mathbf{\Sigma}_T$, is related to the between-level covariance matrix, $\mathbf{\Sigma}_B$, and the within-level covariance matrix, $\mathbf{\Sigma}_W$, via a simple sum: $\mathbf{\Sigma}_T = \mathbf{\Sigma}_B + \mathbf{\Sigma}_W$. In the sample this is denoted $\mathbf{S}_T = \mathbf{S}_B + \mathbf{S}_W$.

Our goal is to show how Gale's derivation scales up to a full correlation matrix, as well as how Muthén's covariance matrices can be standardized into correlation matrices, and that these two solutions are equivalent.

One key mathematical finding in Gale (1987) is how any total correlation r_T can be expressed in terms of r_B and r_W:

$$r_T = \sqrt{\text{ICC}_{x_{1ij}}}\sqrt{\text{ICC}_{x_{2ij}}}\, r_B + \sqrt{1-\text{ICC}_{x_{1ij}}}\sqrt{1-\text{ICC}_{x_{2ij}}}\, r_W \qquad \text{(A.1)}$$

Additionally, Muthén defines $\mathbf{S}_T$, $\mathbf{S}_B$, and $\mathbf{S}_W$ as follows:

$$\mathbf{S}_T = (N-1)^{-1}\sum_{j=1}^{J}\sum_{i=1}^{Nj}\left(x_{ij}-\bar{x}_{..}\right)\left(x_{ij}-\bar{x}_{..}\right)' \qquad \text{(A.2)}$$

$$\mathbf{S}_{PW} = (N-J)^{-1}\sum_{j=1}^{J}\sum_{i=1}^{Nj}\left(x_{ij}-\bar{x}_{.j}\right)\left(x_{ij}-\bar{x}_{.j}\right)' \qquad \text{(A.3)}$$

$$\mathbf{S}_B = (J-1)^{-1}\sum_{j=1}^{J}N_j\left(\bar{x}_{.j}-\bar{x}_{..}\right)\left(\bar{x}_{.j}-\bar{x}_{..}\right)' \qquad \text{(A.4)}$$

Because Muthén uses the matrix formulation, it is more general. We will use Muthén's covariance matrices as a starting point to derive $\mathbf{R_T}$.

To transform a generic covariance matrix $\mathbf{\Sigma}$ to a correlation matrix $\mathbf{R}$, we can pre- and post-multiply $\mathbf{\Sigma}$ by $\mathbf{D}^{-1}$, where $\mathbf{D} = sqrt(diag(\mathbf{\Sigma}))$.

Therefore, in the sample:

$$\mathbf{R}_T = \mathbf{D}^{-1}\mathbf{S}_T\mathbf{D}^{-1} \qquad \text{(A.5)}$$

Substituting:

$$\mathbf{R}_T = \mathbf{D}^{-1}\left(\mathbf{S}_B + \mathbf{S}_W\right)\mathbf{D}^{-1} \tag{A.6}$$

Distributing:

$$\mathbf{R}_T = \mathbf{D}^{-1}\mathbf{S}_B\mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{S}_W\mathbf{D}^{-1} \tag{A.7}$$

Now show what the elements of $\mathbf{R}_T$ will look like after carrying out this multiplication. Then, we can check if those elements match what Gale's math suggests they should be.

$\mathbf{D}$ is a diagonal matrix containing the square roots of the diagonals of $\mathbf{S}_T$. The diagonals of $\mathbf{S}_T$ will be the total variance of each variable. The total variance of each variable is the sum of its between-level and within-level variance, e.g.:

$$\text{var}(x_{1ij}) = \text{var}(x_{1ij})_B + \text{var}(x_{1ij})_W \tag{A.8}$$

Therefore, for two predictors, the matrix $\mathbf{D}^{-1}$ takes the form:

$$\mathbf{D}^{-1} = \begin{bmatrix} \dfrac{1}{\sqrt{\text{var}(x_{1ij})_B + \text{var}(x_{1ij})_W}} & 0 \\ 0 & \dfrac{1}{\sqrt{\text{var}(x_{2ij})_B + \text{var}(x_{2ij})_W}} \end{bmatrix} \tag{A.9}$$

From Muthén's definition, we also know that $\mathbf{S}_B$ will take the following form:

$$\mathbf{S}_B = \begin{bmatrix} \text{var}(x_{1ij})_B & \text{cov}(x_{1ij}, x_{2ij})_B \\ \text{cov}(x_{1ij}, x_{2ij})_B & \text{var}(x_{2ij})_B \end{bmatrix} \tag{A.10}$$

And $\mathbf{S}_W$ will take the following form:

$$\mathbf{S}_W = \begin{bmatrix} \text{var}(x_{1ij})_W & \text{cov}(x_{1ij}, x_{2ij})_W \\ \text{cov}(x_{1ij}, x_{2ij})_W & \text{var}(x_{2ij})_W \end{bmatrix} \tag{A.11}$$

We begin by deriving $\mathbf{D}^{-1}\mathbf{S}_B\mathbf{D}^{-1}$:

$$\mathbf{D}^{-1}\mathbf{S}_B = \begin{bmatrix} \frac{1}{\sqrt{\operatorname{var}(x_{1ij})_B + \operatorname{var}(x_{1ij})_W}} & 0 \\ 0 & \frac{1}{\sqrt{\operatorname{var}(x_{2ij})_B + \operatorname{var}(x_{2ij})_W}} \end{bmatrix} \begin{bmatrix} \operatorname{var}(x_{1ij})_B & \operatorname{cov}(x_{1ij}, x_{2ij})_B \\ \operatorname{cov}(x_{1ij}, x_{2ij})_B & \operatorname{var}(x_{2ij})_B \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\operatorname{var}(x_{1ij})_B}{\sqrt{\operatorname{var}(x_{1ij})_B + \operatorname{var}(x_{1ij})_W}} & \frac{\operatorname{cov}(x_{1ij}, x_{2ij})_B}{\sqrt{\operatorname{var}(x_{1ij})_B + \operatorname{var}(x_{1ij})_W}} \\ \frac{\operatorname{cov}(x_{1ij}, x_{2ij})_B}{\sqrt{\operatorname{var}(x_{2ij})_B + \operatorname{var}(x_{2ij})_W}} & \frac{\operatorname{var}(x_{2ij})_B}{\sqrt{\operatorname{var}(x_{2ij})_B + \operatorname{var}(x_{2ij})_W}} \end{bmatrix}$$

$$\mathbf{D}^{-1}\mathbf{S}_B\mathbf{D}^{-1} = \begin{bmatrix} \frac{\text{var}(x_{1ij})_B}{\sqrt{\text{var}(x_{1ij})_B + \text{var}(x_{1ij})_W}} & \frac{\text{cov}(x_{1ij}, x_{2ij})_B}{\sqrt{\text{var}(x_{1ij})_B + \text{var}(x_{1ij})_W}} \\ \frac{\text{cov}(x_{1ij}, x_{2ij})_B}{\sqrt{\text{var}(x_{2ij})_B + \text{var}(x_{2ij})_W}} & \frac{\text{var}(x_{2ij})_B}{\sqrt{\text{var}(x_{2ij})_B + \text{var}(x_{2ij})_W}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{\text{var}(x_{1ij})_B + \text{var}(x_{1ij})_W}} & 0 \\ 0 & \frac{1}{\sqrt{\text{var}(x_{2ij})_B + \text{var}(x_{2ij})_W}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\text{var}(x_{1ij})_B}{\left(\sqrt{\text{var}(x_{1ij})_B + \text{var}(x_{1ij})_W}\right)\left(\sqrt{\text{var}(x_{1ij})_B + \text{var}(x_{1ij})_W}\right)} & \frac{\text{cov}(x_{1ij}, x_{2ij})_B}{\left(\sqrt{\text{var}(x_{1ij})_B + \text{var}(x_{1ij})_W}\right)\left(\sqrt{\text{var}(x_{2ij})_B + \text{var}(x_{2ij})_W}\right)} \\ \frac{\text{cov}(x_{1ij}, x_{2ij})_B}{\left(\sqrt{\text{var}(x_{1ij})_B + \text{var}(x_{1ij})_W}\right)\left(\sqrt{\text{var}(x_{2ij})_B + \text{var}(x_{2ij})_W}\right)} & \frac{\text{var}(x_{2ij})_B}{\left(\sqrt{\text{var}(x_{2ij})_B + \text{var}(x_{2ij})_W}\right)\left(\sqrt{\text{var}(x_{2ij})_B + \text{var}(x_{2ij})_W}\right)} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\text{var}(x_{1ij})_B}{\text{var}(x_{1ij})_B + \text{var}(x_{1ij})_W} & \frac{\text{cov}(x_{1ij}, x_{2ij})_B}{SD_{x_{1ij}} SD_{x_{2ij}}} \\ \frac{\text{cov}(x_{1ij}, x_{2ij})_B}{SD_{x_{1ij}} SD_{x_{2ij}}} & \frac{\text{var}(x_{2ij})_B}{\text{var}(x_{2ij})_B + \text{var}(x_{2ij})_W} \end{bmatrix}$$

$$= \begin{bmatrix} ICC_{x_{1ij}} & \frac{\text{cov}(x_{1ij}, x_{2ij})_B}{SD_{x_{1ij}} SD_{x_{2ij}}} \\ \frac{\text{cov}(x_{1ij}, x_{2ij})_B}{SD_{x_{1ij}} SD_{x_{2ij}}} & ICC_{x_{2ij}} \end{bmatrix} \quad \text{(A.12)}$$

Next, we derive $\mathbf{D}^{-1}\mathbf{S}_W\mathbf{D}^{-1}$:

$$\mathbf{D}^{-1}\mathbf{S}_W = \begin{bmatrix} \frac{1}{\sqrt{\mathrm{var}(x_{1ij})_B + \mathrm{var}(x_{1ij})_W}} & 0 \\ 0 & \frac{1}{\sqrt{\mathrm{var}(x_{2ij})_B + \mathrm{var}(x_{2ij})_W}} \end{bmatrix} \begin{bmatrix} \mathrm{var}(x_{1ij})_W & \mathrm{cov}(x_{1ij}, x_{2ij})_W \\ \mathrm{cov}(x_{1ij}, x_{2ij})_W & \mathrm{var}(x_{2ij})_W \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\mathrm{var}(x_{1ij})_W}{\sqrt{\mathrm{var}(x_{1ij})_B + \mathrm{var}(x_{1ij})_W}} & \frac{\mathrm{cov}(x_{1ij}, x_{2ij})_W}{\sqrt{\mathrm{var}(x_{1ij})_B + \mathrm{var}(x_{1ij})_W}} \\ \frac{\mathrm{cov}(x_{1ij}, x_{2ij})_W}{\sqrt{\mathrm{var}(x_{2ij})_B + \mathrm{var}(x_{2ij})_W}} & \frac{\mathrm{var}(x_{2ij})_W}{\sqrt{\mathrm{var}(x_{2ij})_B + \mathrm{var}(x_{2ij})_W}} \end{bmatrix}$$

$$\mathbf{D}^{-1}\mathbf{S}_B\mathbf{D}^{-1} = \begin{bmatrix} \frac{\mathrm{var}(x_{1ij})_W}{\sqrt{\mathrm{var}(x_{1ij})_B + \mathrm{var}(x_{1ij})_W}} & \frac{\mathrm{cov}(x_{1ij}, x_{2ij})_W}{\sqrt{\mathrm{var}(x_{1ij})_B + \mathrm{var}(x_{1ij})_W}} \\ \frac{\mathrm{cov}(x_{1ij}, x_{2ij})_W}{\sqrt{\mathrm{var}(x_{2ij})_B + \mathrm{var}(x_{2ij})_W}} & \frac{\mathrm{var}(x_{2ij})_W}{\sqrt{\mathrm{var}(x_{2ij})_B + \mathrm{var}(x_{2ij})_W}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{\mathrm{var}(x_{1ij})_B + \mathrm{var}(x_{1ij})_W}} & 0 \\ 0 & \frac{1}{\sqrt{\mathrm{var}(x_{2ij})_B + \mathrm{var}(x_{2ij})_W}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\mathrm{var}(x_{1ij})_W}{\left(\sqrt{\mathrm{var}(x_{1ij})_B + \mathrm{var}(x_{1ij})_W}\right)\left(\sqrt{\mathrm{var}(x_{1ij})_B + \mathrm{var}(x_{1ij})_W}\right)} & \frac{\mathrm{cov}(x_{1ij}, x_{2ij})_W}{\left(\sqrt{\mathrm{var}(x_{1ij})_B + \mathrm{var}(x_{1ij})_W}\right)\left(\sqrt{\mathrm{var}(x_{2ij})_B + \mathrm{var}(x_{2ij})_W}\right)} \\ \frac{\mathrm{cov}(x_{1ij}, x_{2ij})_W}{\left(\sqrt{\mathrm{var}(x_{1ij})_B + \mathrm{var}(x_{1ij})_W}\right)\left(\sqrt{\mathrm{var}(x_{2ij})_B + \mathrm{var}(x_{2ij})_W}\right)} & \frac{\mathrm{var}(x_{2ij})_W}{\left(\sqrt{\mathrm{var}(x_{2ij})_B + \mathrm{var}(x_{2ij})_W}\right)\left(\sqrt{\mathrm{var}(x_{2ij})_B + \mathrm{var}(x_{2ij})_W}\right)} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\mathrm{var}(x_{1ij})_W}{\mathrm{var}(x_{1ij})_B + \mathrm{var}(x_{1ij})_W} & \frac{\mathrm{cov}(x_{1ij}, x_{2ij})_W}{SD_{x_{1ij}} SD_{x_{2ij}}} \\ \frac{\mathrm{cov}(x_{1ij}, x_{2ij})_W}{SD_{x_{1ij}} SD_{x_{2ij}}} & \frac{\mathrm{var}(x_{2ij})_W}{\mathrm{var}(x_{2ij})_B + \mathrm{var}(x_{2ij})_W} \end{bmatrix}$$

$$= \begin{bmatrix} 1 - ICC_{x_{1ij}} & \frac{\mathrm{cov}(x_{1ij}, x_{2ij})_W}{SD_{x_{1ij}} SD_{x_{2ij}}} \\ \frac{\mathrm{cov}(x_{1ij}, x_{2ij})_W}{SD_{x_{1ij}} SD_{x_{2ij}}} & 1 - ICC_{x_{2ij}} \end{bmatrix} \quad \text{(A.13)}$$

Putting it together:

$$\mathbf{R}_T = \mathbf{D}^{-1}\mathbf{S}_B\mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{S}_W\mathbf{D}^{-1}$$

$$\mathbf{R}_T = \begin{bmatrix} ICC_{x_{1ij}} & \dfrac{\text{cov}(x_{1ij},x_{2ij})_B}{SD_{x_{1ij}}SD_{x_{2ij}}} \\ \dfrac{\text{cov}(x_{1ij},x_{2ij})_B}{SD_{x_{1ij}}SD_{x_{2ij}}} & ICC_{x_{2ij}} \end{bmatrix} + \begin{bmatrix} 1-ICC_{x_{1ij}} & \dfrac{\text{cov}(x_{1ij},x_{2ij})_W}{SD_{x_{1ij}}SD_{x_{2ij}}} \\ \dfrac{\text{cov}(x_{1ij},x_{2ij})_W}{SD_{x_{1ij}}SD_{x_{2ij}}} & 1-ICC_{x_{2ij}} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \dfrac{\text{cov}(x_{1ij},x_{2ij})_B}{SD_{x_{1ij}}SD_{x_{2ij}}} + \dfrac{\text{cov}(x_{1ij},x_{2ij})_W}{SD_{x_{1ij}}SD_{x_{2ij}}} \\ \dfrac{\text{cov}(x_{1ij},x_{2ij})_B}{SD_{x_{1ij}}SD_{x_{2ij}}} + \dfrac{\text{cov}(x_{1ij},x_{2ij})_W}{SD_{x_{1ij}}SD_{x_{2ij}}} & 1 \end{bmatrix} \tag{A.14}$$

Using Muthén's covariance matrices as a starting point, we have shown what the off-diagonals of $\mathbf{R}_T$ will contain. Now, we can check whether this solution matches Gale's formula for a single correlation coefficient, r_T .

We need to show that

$$\underbrace{\sqrt{\text{ICC}_{x_{1ij}}}\sqrt{\text{ICC}_{x_{2ij}}}\,r_B + \sqrt{1-\text{ICC}_{x_{1ij}}}\sqrt{1-\text{ICC}_{x_{2ij}}}\,r_W}_{\text{from Gale}} = \underbrace{\frac{\text{cov}(x_{1ij},x_{2ij})_B}{SD_{x_{1ij}}SD_{x_{2ij}}} + \frac{\text{cov}(x_{1ij},x_{2ij})_W}{SD_{x_{1ij}}SD_{x_{2ij}}}}_{\text{from Muthen}}$$

Beginning with the first piece of each equation, we can show that

$$\sqrt{\text{ICC}_{x_{1ij}}}\sqrt{\text{ICC}_{x_{2ij}}}\,r_B = \frac{\text{cov}(x_{1ij},x_{2ij})_B}{SD_{x_{1ij}}SD_{x_{2ij}}}$$

Recall from Eq. (A.4) that Muthén defines the between-level covariance matrix as:

$$\mathbf{S}_B = (J-1)^{-1}\sum_{j=1}^{J} N_j\left(\bar{x}_{.j} - \bar{x}_{..}\right)\left(\bar{x}_{.j} - \bar{x}_{..}\right)'$$

Applying this definition to $\text{cov}(x_{1ij},x_{2ij})$, we obtain:

$$\text{cov}(x_{1ij},x_{2ij})_B = \frac{\sum_{j=1}^{J} N_j\left(\bar{x}_{1.j} - \bar{x}_{1..}\right)\left(\bar{x}_{2.j} - \bar{x}_{..}\right)}{(J-1)} \tag{A.15}$$

To convert $\text{cov}(x_{1ij}, x_{2ij})$ into r_B, we standardize by $\sqrt{\text{var}(x_{1ij})_B}\sqrt{\text{var}(x_{2ij})_B}$, and this matches Gale's definition of r_B:

$$r_B = \frac{\sum_{j=1}^{J} N_j\left(\bar{x}_{1.j} - \bar{x}_{1..}\right)\left(\bar{x}_{2.j} - \bar{x}_{..}\right)}{(J-1)\sqrt{\text{var}(x_{1ij})_B}\sqrt{\text{var}(x_{2ij})_B}} \tag{A.16}$$

We can also rewrite $\sqrt{\text{ICC}_{x_{1ij}}}\sqrt{\text{ICC}_{x_{2ij}}}$ as follows:

$$\sqrt{\text{ICC}_{x_{1ij}}} = \sqrt{\frac{\text{var}(x_{1ij})_B}{\text{var}(x_{1ij})_B + \text{var}(x_{1ij})_W}} = \frac{\sqrt{\text{var}(x_{1ij})_B}}{\sqrt{\text{var}(x_{1ij})_B + \text{var}(x_{1ij})_W}}$$

(A.17, A.18)

$$\sqrt{\text{ICC}_{x_{2ij}}} = \sqrt{\frac{\text{var}(x_{2ij})_B}{\text{var}(x_{2ij})_B + \text{var}(x_{2ij})_W}} = \frac{\sqrt{\text{var}(x_{2ij})_B}}{\sqrt{\text{var}(x_{2ij})_B + \text{var}(x_{2ij})_W}}$$

With everything re-expressed, we can now re-express Gale's formula:

$$\sqrt{\text{ICC}_{x_{1ij}}}\sqrt{\text{ICC}_{x_{2ij}}}\, r_B = \left(\frac{\sqrt{\text{var}(x_{1ij})_B}}{\sqrt{\text{var}(x_{1ij})_B + \text{var}(x_{1ij})_W}}\right)\left(\frac{\sqrt{\text{var}(x_{2ij})_B}}{\sqrt{\text{var}(x_{2ij})_B + \text{var}(x_{2ij})_W}}\right)\left(\frac{\sum_{j=1}^{J} N_j\left(\bar{x}_{1.j} - \bar{x}_{1..}\right)\left(\bar{x}_{2.j} - \bar{x}_{..}\right)}{(J-1)\sqrt{\text{var}(x_{1ij})_B}\sqrt{\text{var}(x_{2ij})_B}}\right)$$

Cancelling the red terms:

$$\left(\frac{\color{red}{\sqrt{\text{var}(x_{1ij})_B}}}{\sqrt{\text{var}(x_{1ij})_B + \text{var}(x_{1ij})_W}}\right)\left(\frac{\color{red}{\sqrt{\text{var}(x_{2ij})_B}}}{\sqrt{\text{var}(x_{2ij})_B + \text{var}(x_{2ij})_W}}\right)\left(\frac{\sum_{j=1}^{J} N_j\left(\bar{x}_{1.j} - \bar{x}_{1..}\right)\left(\bar{x}_{2.j} - \bar{x}_{..}\right)}{(J-1)\color{red}{\sqrt{\text{var}(x_{1ij})_B}\sqrt{\text{var}(x_{2ij})_B}}}\right)$$

$$= \frac{\sum_{j=1}^{J} N_j\left(\bar{x}_{1.j} - \bar{x}_{1..}\right)\left(\bar{x}_{2.j} - \bar{x}_{..}\right)}{(J-1)\sqrt{\text{var}(x_{1ij})_B + \text{var}(x_{1ij})_W}\sqrt{\text{var}(x_{2ij})_B + \text{var}(x_{2ij})_W}}$$

$$= \frac{\text{cov}(x_{1ij}, x_{2ij})_B}{SD_{x_{1ij}} SD_{x_{2ij}}} \tag{A.19}$$

Which matches what we derived from the covariance matrix formulation.

Following the same steps, we can show that $\sqrt{1-\mathrm{ICC}_{x_{1ij}}}\sqrt{1-\mathrm{ICC}_{x_{2ij}}}r_W = \dfrac{\mathrm{cov}(x_{1ij}, x_{2ij})_W}{SD_{x_{1ij}} SD_{x_{2ij}}}$

$$\sqrt{1-\mathrm{ICC}_{x_{1ij}}}\sqrt{1-\mathrm{ICC}_{x_{2ij}}}r_W$$

$$= \left(\frac{\sqrt{\mathrm{var}(x_{1ij})_W}}{\sqrt{\mathrm{var}(x_{1ij})_B + \mathrm{var}(x_{1ij})_W}}\right)\left(\frac{\sqrt{\mathrm{var}(x_{2ij})_W}}{\sqrt{\mathrm{var}(x_{2ij})_B + \mathrm{var}(x_{2ij})_W}}\right)\left(\frac{\sum_{j=1}^{J}\sum_{i=1}^{Nj}\left(x_{1ij}-\bar{x}_{1.j}\right)\left(x_{2ij}-\bar{x}_{2.j}\right)}{(N-J)\sqrt{\mathrm{var}(x_{1ij})_W}\sqrt{\mathrm{var}(x_{2ij})_W}}\right)$$

$$= \left(\frac{\sqrt{\mathrm{var}(x_{1ij})_W}}{\sqrt{\mathrm{var}(x_{1ij})_B + \mathrm{var}(x_{1ij})_W}}\right)\left(\frac{\sqrt{\mathrm{var}(x_{2ij})_W}}{\sqrt{\mathrm{var}(x_{2ij})_B + \mathrm{var}(x_{2ij})_W}}\right)\left(\frac{\sum_{j=1}^{J}\sum_{i=1}^{Nj}\left(x_{1ij}-\bar{x}_{1.j}\right)\left(x_{2ij}-\bar{x}_{2.j}\right)}{(N-J)\sqrt{\mathrm{var}(x_{1ij})_W}\sqrt{\mathrm{var}(x_{2ij})_W}}\right)$$

$$= \frac{\sum_{j=1}^{J}\sum_{i=1}^{Nj}\left(x_{1ij}-\bar{x}_{1.j}\right)\left(x_{2ij}-\bar{x}_{2.j}\right)}{(N-J)\sqrt{\mathrm{var}(x_{1ij})_B + \mathrm{var}(x_{1ij})_W}\sqrt{\mathrm{var}(x_{2ij})_B + \mathrm{var}(x_{2ij})_W}}$$

$$= \frac{\mathrm{cov}(x_{1ij}, x_{2ij})_W}{SD_{x_{1ij}} SD_{x_{2ij}}} \tag{A.20}$$

Which matches what we derived from the covariance formulation.

Appendix B: Deriving the maximally general form of the GLS estimator

Beginning with Scott and Holt's (1982) original equation:

$$\hat{\beta}_{GLS} = \left\{ \sum_{j=1}^{J} \frac{X_{Bj}^T X_{Bj}}{1+(n_j-1)\rho} + \sum_{j=1}^{J} \frac{X_{Wj}^T X_{Wj}}{1-\rho} \right\}^{-1} \times \left\{ \sum_{j=1}^{J} \frac{X_{Bj}^T Y_j}{1+(n_j-1)\rho} + \sum_{j=1}^{J} \frac{X_{Wj}^T Y_j}{1-\rho} \right\} \qquad \text{(B.1)}$$

Re-expressing X_{Bj} in matrix notation:

$$\begin{aligned} X_{Bj} &= n_j^{-1}\mathbf{1}_{n_j}\mathbf{1}'_{n_j}\left[\mathbf{1}_{n_j} \mid \mathbf{X}_j\right] \\ &= n_j^{-1}\left[\mathbf{1}_{n_j}\underbrace{\mathbf{1}'_{n_j}\mathbf{1}_{n_j}}_{n_j} \;\Big|\; \mathbf{1}_{n_j}\underbrace{\mathbf{1}'_{n_j}\mathbf{X}_j}_{\substack{\text{col sums}\\ \text{of } \mathbf{X}_j}}\right] \\ &= n_j^{-1}\left[\mathbf{1}_{n_j} n_j \mid \mathbf{1}_{n_j}\mathbf{x}'_j\right] \\ &= \left[\mathbf{1}_{n_j} \mid n_j^{-1}\mathbf{1}_{n_j}\mathbf{x}'_j\right] \end{aligned} \qquad \text{(B.2)}$$

Where $\mathbf{X}_j$ contains all the predictor values for cluster *j* and $\mathbf{x}'_j$ is a row vector of cluster sums for the *k* predictors. Thus, we can re-express the sums in $\mathbf{x}'_j$ as n_j times the cluster means, or $n_j\bar{\mathbf{x}}'_j$:

$$\begin{aligned} X_{Bj} &= \left[\mathbf{1}_{n_j} \mid n_j^{-1}\mathbf{1}_{n_j}\mathbf{x}'_j\right] \\ &= \left[\mathbf{1}_{n_j} \mid n_j^{-1}\mathbf{1}_{n_j} n_j\bar{\mathbf{x}}'_j\right] \\ &= \left[\mathbf{1}_{n_j} \mid \mathbf{1}_{n_j}\bar{\mathbf{x}}'_j\right] \end{aligned} \qquad \text{(B.3)}$$

Following a similar process to express X_{Wj} in matrix notation:

$$\begin{aligned} X_{Wj} &= \left(\mathbf{I}_{n_j} - n_j^{-1}\mathbf{1}_{n_j}\mathbf{1}'_{n_j}\right)\left[\mathbf{1}_{n_j} \mid \mathbf{X}_j\right] \\ &= \mathbf{I}_{n_j}\left[\mathbf{1}_{n_j} \mid \mathbf{X}_j\right] - n_j^{-1}\mathbf{1}_{n_j}\mathbf{1}'_{n_j}\left[\mathbf{1}_{n_j} \mid \mathbf{X}_j\right] \\ &= \left[\mathbf{1}_{n_j} \mid \mathbf{X}_j\right] - \left[\mathbf{1}_{n_j} \mid \mathbf{1}_{n_j}\bar{\mathbf{x}}'_j\right] \\ &= \left[\mathbf{0}_{n_j} \mid \mathbf{X}_j - \mathbf{1}_{n_j}\bar{\mathbf{x}}'_j\right] \end{aligned} \qquad \text{(B.4)}$$

Then, expressing $X_{Bj}^T X_{Bj}$ in matrix notation:

$$
\begin{aligned}
X_{Bj}^T X_{Bj} &= \left[\mathbf{1}_{n_j} \;\middle|\; \mathbf{1}_{n_j}\bar{\mathbf{x}}_j'\right]' \left[\mathbf{1}_{n_j} \;\middle|\; \mathbf{1}_{n_j}\bar{\mathbf{x}}_j'\right] \\
&= \left[\begin{array}{c} \mathbf{1}_{n_j}' \\ \hdashline \bar{\mathbf{x}}_j \mathbf{1}_{n_j}' \end{array}\right] \left[\mathbf{1}_{n_j} \;\middle|\; \mathbf{1}_{n_j}\bar{\mathbf{x}}_j'\right] \\
&= \left[\begin{array}{c:c} \mathbf{1}_{n_j}'\mathbf{1}_{n_j} & \mathbf{1}_{n_j}'\mathbf{1}_{n_j}\bar{\mathbf{x}}_j' \\ \hdashline \bar{\mathbf{x}}_j\mathbf{1}_{n_j}'\mathbf{1}_{n_j} & \bar{\mathbf{x}}_j\mathbf{1}_{n_j}'\mathbf{1}_{n_j}\bar{\mathbf{x}}_j' \end{array}\right] \\
&= \left[\begin{array}{c:c} n_j & n_j\bar{\mathbf{x}}_j' \\ \hdashline n_j\bar{\mathbf{x}}_j & n_j\bar{\mathbf{x}}_j\bar{\mathbf{x}}_j' \end{array}\right]
\end{aligned}
\tag{B.5}
$$

And expressing $X_{Wj}^T X_{Wj}$ in matrix notation:

$$
\begin{aligned}
X_{Wj}^T X_{Wj} &= \left[\mathbf{0}_{n_j} \;\middle|\; \mathbf{X}_j - \mathbf{1}_{n_j}\bar{\mathbf{x}}_j'\right]' \left[\mathbf{0}_{n_j} \;\middle|\; \mathbf{X}_j - \mathbf{1}_{n_j}\bar{\mathbf{x}}_j'\right] \\
&= \left[\begin{array}{c} \mathbf{0}_{n_j}' \\ \hdashline \mathbf{X}_j' - \bar{\mathbf{x}}_j\mathbf{1}_{n_j}' \end{array}\right] \left[\mathbf{0}_{n_j} \;\middle|\; \mathbf{X}_j - \mathbf{1}_{n_j}\bar{\mathbf{x}}_j'\right] \\
&= \left[\begin{array}{c:c} \mathbf{0}_{n_j}'\mathbf{0}_{n_j} & \mathbf{0}_{n_j}'\left(\mathbf{X}_j - \mathbf{1}_{n_j}\bar{\mathbf{x}}_j'\right) \\ \hdashline \left(\mathbf{X}_j' - \bar{\mathbf{x}}_j\mathbf{1}_{n_j}'\right)\mathbf{0}_{n_j} & \left(\mathbf{X}_j' - \bar{\mathbf{x}}_j\mathbf{1}_{n_j}'\right)\left(\mathbf{X}_j - \mathbf{1}_{n_j}\bar{\mathbf{x}}_j'\right) \end{array}\right] \\
&= \left[\begin{array}{c:c} 0 & \mathbf{0}_{n_j}' \\ \hdashline \mathbf{0}_{n_j} & \left(\mathbf{X}_j'\mathbf{X}_j - \mathbf{X}_j'\mathbf{1}_{n_j}\bar{\mathbf{x}}_j' - \bar{\mathbf{x}}_j\mathbf{1}_{n_j}'\mathbf{X}_j + \bar{\mathbf{x}}_j\mathbf{1}_{n_j}'\mathbf{1}_{n_j}\bar{\mathbf{x}}_j'\right) \end{array}\right] \\
&= \left[\begin{array}{c:c} 0 & \mathbf{0}_{n_j}' \\ \hdashline \mathbf{0}_{n_j} & \left(\mathbf{X}_j'\mathbf{X}_j - \mathbf{X}_j'\mathbf{1}_{n_j}\bar{\mathbf{x}}_j' - \bar{\mathbf{x}}_j\mathbf{1}_{n_j}'\mathbf{X}_j + n_j\bar{\mathbf{x}}_j\bar{\mathbf{x}}_j'\right) \end{array}\right]
\end{aligned}
\tag{B.6}
$$

In the 2,2 cell, notice that $\mathbf{X}'_j\mathbf{1}_{n_j}$ is a column vector of predictor sums in cluster *j*, which is the same as n_j times a vector of cluster means of the predictors, or $n_j\bar{\mathbf{x}}_j$. Therefore (B.6) can be simplified:

$$X_{Wj}^T X_{Wj} = \left[\begin{array}{c|c} 0 & \mathbf{0}'_{n_j} \\ \hline \mathbf{0}_{n_j} & \left(\mathbf{X}'_j\mathbf{X}_j - \mathbf{X}'_j\mathbf{1}_{n_j}\bar{\mathbf{x}}'_j - \bar{\mathbf{x}}_j\mathbf{1}'_{n_j}\mathbf{X}_j + n_j\bar{\mathbf{x}}_j\bar{\mathbf{x}}'_j\right) \end{array}\right]$$

$$= \left[\begin{array}{c|c} 0 & \mathbf{0}'_{n_j} \\ \hline \mathbf{0}_{n_j} & \left(\mathbf{X}'_j\mathbf{X}_j - n_j\bar{\mathbf{x}}_j\bar{\mathbf{x}}'_j - n_j\bar{\mathbf{x}}_j\bar{\mathbf{x}}'_j + n_j\bar{\mathbf{x}}_j\bar{\mathbf{x}}'_j\right) \end{array}\right] \qquad \text{(B.7)}$$

$$= \left[\begin{array}{c|c} 0 & \mathbf{0}'_{n_j} \\ \hline \mathbf{0}_{n_j} & \left(\mathbf{X}'_j\mathbf{X}_j - n_j\bar{\mathbf{x}}_j\bar{\mathbf{x}}'_j\right) \end{array}\right]$$

Inserting (B.5) and (B.7) into the first term of (B.1):

$$\left\{\sum_{j=1}^{J}\frac{X_{Bj}^T X_{Bj}}{1+\left(n_j-1\right)\rho} + \sum_{j=1}^{J}\frac{X_{Wj}^T X_{Wj}}{1-\rho}\right\}^{-1} = \left\{\sum_{j=1}^{J}\left(\frac{X_{Bj}^T X_{Bj}}{1+\left(n_j-1\right)\rho} + \frac{X_{Wj}^T X_{Wj}}{1-\rho}\right)\right\}^{-1} \qquad \text{(B.8)}$$

$$= \left\{\sum_{j=1}^{J}\left(\left(1+\left(n_j-1\right)\rho\right)^{-1} X_{Bj}^T X_{Bj} + \left(1-\rho\right)^{-1} X_{Wj}^T X_{Wj}\right)\right\}^{-1}$$

$$= \left\{\sum_{j=1}^{J}\left(\left(1+\left(n_j-1\right)\rho\right)^{-1}\left[\begin{array}{c|c} n_j & n_j\bar{\mathbf{x}}'_j \\ \hline n_j\bar{\mathbf{x}}_j & n_j\bar{\mathbf{x}}_j\bar{\mathbf{x}}'_j \end{array}\right] + \left(1-\rho\right)^{-1}\left[\begin{array}{c|c} 0 & \mathbf{0}'_{n_j} \\ \hline \mathbf{0}_{n_j} & \left(\mathbf{X}'_j\mathbf{X}_j - n_j\bar{\mathbf{x}}_j\bar{\mathbf{x}}'_j\right) \end{array}\right]\right)\right\}^{-1}$$

$$= \left\{\sum_{j=1}^{J}\left[\begin{array}{c|c} \left(1+\left(n_j-1\right)\rho\right)^{-1} n_j & \left(1+\left(n_j-1\right)\rho\right)^{-1} n_j\bar{\mathbf{x}}'_j \\ \hline \left(1+\left(n_j-1\right)\rho\right)^{-1} n_j\bar{\mathbf{x}}_j & \left(1+\left(n_j-1\right)\rho\right)^{-1} n_j\bar{\mathbf{x}}_j\bar{\mathbf{x}}'_j + \left(1-\rho\right)^{-1}\left(\mathbf{X}'_j\mathbf{X}_j - n_j\bar{\mathbf{x}}_j\bar{\mathbf{x}}'_j\right) \end{array}\right]\right\}^{-1}$$

Next, we express the second term of (B.1) in matrix notation, from earlier, we know

$X_{Bj} = \left[\mathbf{1}_{n_j} \mid \mathbf{1}_{n_j}\bar{\mathbf{x}}'_j\right]$ and $X_{Wj} = \left[\mathbf{0}_{n_j} \mid \mathbf{X}_j - \mathbf{1}_{n_j}\bar{\mathbf{x}}'_j\right]$. Inserting these into into the second term of (B.1), we obtain:

$$\left\{\sum_{j=1}^{J}\left(\left(1+(n_j-1)\rho\right)^{-1}\left[\mathbf{1}_{n_j} \mid \mathbf{1}_{n_j}\bar{\mathbf{x}}'_j\right]^T Y_j + \left(1-\rho\right)^{-1}\left[\mathbf{0}_{n_j} \mid \mathbf{X}_j - \mathbf{1}_{n_j}\bar{\mathbf{x}}'_j\right]^T Y_j\right)\right\} \qquad \text{(B.9)}$$

Putting together (B.8) and (B.9), the general form of (B.1) in matrix notation is:

$$\hat{\beta}_{GLS} = \left\{ \sum_{j=1}^{J} \left[\begin{array}{c|c} \left(1+\left(n_j-1\right)\rho\right)^{-1} n_j & \left(1+\left(n_j-1\right)\rho\right)^{-1} n_j \overline{\mathbf{x}}_j' \\ \hline \left(1+\left(n_j-1\right)\rho\right)^{-1} n_j \overline{\mathbf{x}}_j & \left(1+\left(n_j-1\right)\rho\right)^{-1} n_j \overline{\mathbf{x}}_j \overline{\mathbf{x}}_j' + \left(1-\rho\right)^{-1} \left(\mathbf{X}_j' \mathbf{X}_j - n_j \overline{\mathbf{x}}_j \overline{\mathbf{x}}_j' \right) \end{array} \right] \right\}^{-1}$$
$$\times \left\{ \sum_{j=1}^{J} \left(\left(1+(n_j-1)\rho\right)^{-1} \left[\mathbf{1}_{n_j} \mid \mathbf{1}_{n_j} \overline{\mathbf{x}}_j' \right]^T Y_j + \left(1-\rho\right)^{-1} \left[\mathbf{0}_{n_j} \mid \mathbf{X}_j - \mathbf{1}_{n_j} \overline{\mathbf{x}}_j' \right]^T Y_j \right) \right\}$$

(B.10)

Appendix C: Expressing the maximally general form of the GLS estimator in terms of within- and between-cluster predictor covariance matrices

Our goal is to reframe the GLS formula in terms of covariance matrices rather than sums of cross-products to demonstrate how the *between* and *within* parts of predictors inform conflated GLS estimates. To do this, we use Muthén's (1990) formulas for the *between* and *pooled-within* covariance matrices. A complicating factor is that Muthén's matrices are for endogenous variables **y**, whereas our GLS formula is in terms of **x** with 1 in the first slot as a multiplier for the intercept.

The *between* part in the GLS formula is (after distributing the summation and assuming equal cluster sizes for simplicity):

$$\begin{aligned}\mathbf{Q}_B &= \sum_{j=1}^{J}\left(1+\left(n_j-1\right)\rho\right)^{-1} n_j \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j' \\ &= \left(1+\left(n-1\right)\rho\right)^{-1} n \sum_{j=1}^{J} \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j'\end{aligned} \tag{C.1}$$

We want to express $\mathbf{Q}_B$ in terms of Muthén's $\mathbf{S}_B$. Muthén's set of engodenous variables has no 1 in the first slot. Returning to $\mathbf{Q}_B$ and letting $\bar{\mathbf{z}}_j$ be the part of $\bar{\mathbf{x}}_j$ excluding the 1 in the first element, re-express the sum of cross-products as:

$$\sum_{j=1}^{J} \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j' = \sum_{j=1}^{J} \begin{bmatrix} 1 & \bar{\mathbf{z}}_j' \\ \bar{\mathbf{z}}_j & \bar{\mathbf{z}}_j \bar{\mathbf{z}}_j' \end{bmatrix} \tag{C.2}$$

Yielding:

$$\mathbf{Q}_B = \left(1+\left(n-1\right)\rho\right)^{-1} n \sum_{j=1}^{J} \begin{bmatrix} 1 & \bar{\mathbf{z}}_j' \\ \bar{\mathbf{z}}_j & \bar{\mathbf{z}}_j \bar{\mathbf{z}}_j' \end{bmatrix} \tag{C.3}$$

Now we focus our attention on $\bar{\mathbf{z}}_j \bar{\mathbf{z}}_j'$. After assuming equal cluster sizes and grand mean centering, Muthén's formula is:

$$\begin{aligned}\mathbf{S}_B &= \left(J-1\right)^{-1} \sum_{j=1}^{J} n_j \bar{\mathbf{z}}_j \bar{\mathbf{z}}_j' \\ &= \left(J-1\right)^{-1} n \sum_{j=1}^{J} \bar{\mathbf{z}}_j \bar{\mathbf{z}}_j'\end{aligned} \tag{C.4}$$

Substituting, we have:

$$\mathbf{Q}_B = \left(1+(n-1)\rho\right)^{-1} n\sum_{j=1}^{J}\begin{bmatrix} 1 & \overline{\mathbf{z}}'_j \\ \overline{\mathbf{z}}_j & \overline{\mathbf{z}}_j\overline{\mathbf{z}}'_j \end{bmatrix}$$

$$= \left(1+(n-1)\rho\right)^{-1} n\begin{bmatrix} \sum_{j=1}^{J} 1 & \sum_{j=1}^{J}\overline{\mathbf{z}}'_j \\ \sum_{j=1}^{J}\overline{\mathbf{z}}_j & \sum_{j=1}^{J}\overline{\mathbf{z}}_j\overline{\mathbf{z}}'_j \end{bmatrix}$$

$$= \left(1+(n-1)\rho\right)^{-1} n\begin{bmatrix} J & \sum_{j=1}^{J}\overline{\mathbf{z}}'_j \\ \sum_{j=1}^{J}\overline{\mathbf{z}}_j & (J-1)n^{-1}\mathbf{S}_B \end{bmatrix}$$

$$= \left(1+(n-1)\rho\right)^{-1}\begin{bmatrix} Jn & n\sum_{j=1}^{J}\overline{\mathbf{z}}'_j \\ n\sum_{j=1}^{J}\overline{\mathbf{z}}_j & (J-1)\mathbf{S}_B \end{bmatrix} \tag{C.5}$$

Thus, we have shown that $\mathbf{Q}_B$ from the maximally general GLS estimator can be expressed in terms of the between-cluster covariance matrix of predictors.

The *within* part in the GLS formula is:

$$\mathbf{Q}_W = \sum_{j=1}^{J}(1-\rho)^{-1}\left(\mathbf{X}'_j\mathbf{X}_j - n_j\overline{\mathbf{x}}_j\overline{\mathbf{x}}'_j\right)$$
$$= (1-\rho)^{-1}\sum_{j=1}^{J}\left(\mathbf{X}'_j\mathbf{X}_j - n\overline{\mathbf{x}}_j\overline{\mathbf{x}}'_j\right) \tag{C.6}$$

We want to express $\mathbf{Q}_W$ in terms of Muthén's $\mathbf{S}_{PW}$. Now also letting $\mathbf{Z}_j$ represent the part of $\mathbf{X}_j$ excluding the 1 in the first element, we can re-express the sums of cross-products as:

$$\begin{aligned}
\sum_{j=1}^{J}\left(\mathbf{X}_j'\mathbf{X}_j - n\overline{\mathbf{x}}_j\overline{\mathbf{x}}_j'\right) &= \sum_{j=1}^{J}\mathbf{X}_j'\mathbf{X}_j - n\sum_{j=1}^{J}\overline{\mathbf{x}}_j\overline{\mathbf{x}}_j' \\
&= \sum_{j=1}^{J}\begin{bmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\mathbf{Z}_j \\ \mathbf{Z}_j'\mathbf{1} & \mathbf{Z}_j'\mathbf{Z}_j \end{bmatrix} - n\sum_{j=1}^{J}\begin{bmatrix} 1 & \overline{\mathbf{z}}_j' \\ \overline{\mathbf{z}}_j & \overline{\mathbf{z}}_j\overline{\mathbf{z}}_j' \end{bmatrix} \\
&= \sum_{j=1}^{J}\begin{bmatrix} n & n\overline{\mathbf{z}}' \\ n\overline{\mathbf{z}}_j & \mathbf{Z}_j'\mathbf{Z}_j \end{bmatrix} - n\sum_{j=1}^{J}\begin{bmatrix} 1 & \overline{\mathbf{z}}_j' \\ \overline{\mathbf{z}}_j & \overline{\mathbf{z}}_j\overline{\mathbf{z}}_j' \end{bmatrix} \\
&= \sum_{j=1}^{J}\begin{bmatrix} 0 & n\overline{\mathbf{z}}' - n\overline{\mathbf{z}}_j' \\ n\overline{\mathbf{z}}_j - n\overline{\mathbf{z}}_j & \mathbf{Z}_j'\mathbf{Z}_j - n\overline{\mathbf{z}}_j\overline{\mathbf{z}}_j' \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 \\ 0 & \sum_{j=1}^{J}\left(\mathbf{Z}_j'\mathbf{Z}_j - n\overline{\mathbf{z}}_j\overline{\mathbf{z}}_j'\right) \end{bmatrix}
\end{aligned} \tag{C.7}$$

After assuming equal cluster sizes and that predictors have been grand mean centered, Muthén's formula is:

$$\begin{aligned}
\mathbf{S}_{PW} &= \left(N-J\right)^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n_j}\left(\mathbf{z}_{ij} - \overline{\mathbf{z}}_{.j}\right)\left(\mathbf{z}_{ij} - \overline{\mathbf{z}}_{.j}\right)' \\
&= \left(N-J\right)^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n}\left(\mathbf{z}_{ij} - \overline{\mathbf{z}}_{.j}\right)\left(\mathbf{z}_{ij} - \overline{\mathbf{z}}_{.j}\right)'
\end{aligned} \tag{C.8}$$

To be able to substitute terms from our re-expression of $\mathbf{Q}_W$ with Muthén's $\mathbf{S}_{PW}$, we need to rearrange the cross-product in $\mathbf{S}_{PW}$:

$$\begin{aligned}
\sum_{j=1}^{J}\sum_{i=1}^{n}\left(\mathbf{z}_{ij} - \overline{\mathbf{z}}_j\right)\left(\mathbf{z}_{ij} - \overline{\mathbf{z}}_j\right)' &= \sum_{j=1}^{J}\sum_{i=1}^{n}\left(\mathbf{z}_{ij}\mathbf{z}'_{ij} - \mathbf{z}_{ij}\overline{\mathbf{z}}_j' - \overline{\mathbf{z}}_j\mathbf{z}'_{ij} + \overline{\mathbf{z}}_j\overline{\mathbf{z}}_j'\right) \\
&= \sum_{j=1}^{J}\sum_{i=1}^{n}\mathbf{z}_{ij}\mathbf{z}'_{ij} - \sum_{j=1}^{J}\sum_{i=1}^{n}\mathbf{z}_{ij}\overline{\mathbf{z}}_j' - \sum_{j=1}^{J}\sum_{i=1}^{n}\overline{\mathbf{z}}_j\mathbf{z}'_{ij} + \sum_{j=1}^{J}\sum_{i=1}^{n}\overline{\mathbf{z}}_j\overline{\mathbf{z}}_j' \\
&= \sum_{j=1}^{J}\sum_{i=1}^{n}\mathbf{z}_{ij}\mathbf{z}'_{ij} - \sum_{j=1}^{J}\left(\sum_{i=1}^{n}\mathbf{z}_{ij}\right)\overline{\mathbf{z}}_j' - \sum_{j=1}^{J}\overline{\mathbf{z}}_j\sum_{i=1}^{n}\mathbf{z}'_{ij} + n\sum_{j=1}^{J}\overline{\mathbf{z}}_j\overline{\mathbf{z}}_j' \\
&= \sum_{j=1}^{J}\sum_{i=1}^{n}\mathbf{z}_{ij}\mathbf{z}'_{ij} - n\sum_{j=1}^{J}\left(n^{-1}\sum_{i=1}^{n}\mathbf{z}_{ij}\right)\overline{\mathbf{z}}_j' - n\sum_{j=1}^{J}\overline{\mathbf{z}}_j n^{-1}\sum_{i=1}^{n}\mathbf{z}'_{ij} + n\sum_{j=1}^{J}\overline{\mathbf{z}}_j\overline{\mathbf{z}}_j'
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{J}\sum_{i=1}^{n}\mathbf{z}_{ij}\mathbf{z}'_{ij} - n\sum_{j=1}^{J}\overline{\mathbf{z}}_{j}\overline{\mathbf{z}}'_{j} - n\sum_{j=1}^{J}\overline{\mathbf{z}}_{j}\overline{\mathbf{z}}'_{j} + n\sum_{j=1}^{J}\overline{\mathbf{z}}_{j}\overline{\mathbf{z}}'_{j} \\
&= \sum_{j=1}^{J}\sum_{i=1}^{n}\mathbf{z}_{ij}\mathbf{z}'_{ij} - n\sum_{j=1}^{J}\overline{\mathbf{z}}_{j}\overline{\mathbf{z}}'_{j}
\end{aligned}
\tag{C.9}
$$

Recall that the cross-product from $\mathbf{Q}_W$, expressed in the final line of C.7, is:

$$
\sum_{j=1}^{J}\left(\mathbf{Z}'_j\mathbf{Z}_j - n\overline{\mathbf{z}}_j\overline{\mathbf{z}}'_j\right) = \sum_{j=1}^{J}\mathbf{Z}'_j\mathbf{Z}_j - n\sum_{j=1}^{J}\overline{\mathbf{z}}_j\overline{\mathbf{z}}'_j \tag{C.10}
$$

The second term of (C.10) matches the second term in the final line of (C.9). However, the first terms do not match yet. Now we must show that $\mathbf{Z}'_j\mathbf{Z}_j = \sum_{i=1}^{n}\mathbf{z}_{ij}\mathbf{z}'_{ij}$.

Express $\mathbf{Z}_j$ as: $\underset{n\times p}{\mathbf{Z}_j} = \begin{bmatrix} \mathbf{z}'_{ij|i=1} \\ \mathbf{z}'_{ij|i=2} \\ \vdots \\ \mathbf{z}'_{ij|i=n} \end{bmatrix}$ where each $\mathbf{z}'_{ij}$ is $1\times p$, where p is the number of predictors.

Thus, $\underset{p\times n}{\mathbf{Z}'_j} = \begin{bmatrix} \mathbf{z}_{ij|i=1} & \mathbf{z}_{ij|i=2} & \cdots & \mathbf{z}_{ij|i=n} \end{bmatrix}$

and $\mathbf{Z}'_j\mathbf{Z}_j = \sum_{i=1}^{n}\mathbf{z}_{ij}\mathbf{z}'_{ij}$

We just showed that

$$
\begin{aligned}
\mathbf{S}_{PW} &= (N-J)^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n_j}\left(\mathbf{z}_{ij} - \overline{\mathbf{z}}_{.j}\right)\left(\mathbf{z}_{ij} - \overline{\mathbf{z}}_{.j}\right)' \\
&= (N-J)^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n}\left(\mathbf{z}_{ij} - \overline{\mathbf{z}}_{.j}\right)\left(\mathbf{z}_{ij} - \overline{\mathbf{z}}_{.j}\right)' \\
&= (N-J)^{-1}\sum_{j=1}^{J}\left(\mathbf{Z}'_j\mathbf{Z}_j - n\overline{\mathbf{z}}_j\overline{\mathbf{z}}'_j\right)
\end{aligned}
\tag{C.11}
$$

Thus,

$$\begin{aligned}
\mathbf{Q}_W &= (1-\rho)^{-1}\sum_{j=1}^{J}\left(\mathbf{X}_j'\mathbf{X}_j - n\bar{\mathbf{x}}_j\bar{\mathbf{x}}_j'\right) \\
&= (1-\rho)^{-1}\begin{bmatrix} 0 & \sum_{j=1}^{J}\left(\mathbf{Z}_j - n\bar{\mathbf{z}}_j'\right) \\ \sum_{j=1}^{J}\left(\mathbf{Z}_j' - n\bar{\mathbf{z}}_j\right) & \sum_{j=1}^{J}\left(\mathbf{Z}_j'\mathbf{Z}_j - n\bar{\mathbf{z}}_j\bar{\mathbf{z}}_j'\right) \end{bmatrix} \\
&= (1-\rho)^{-1}\begin{bmatrix} 0 & \sum_{j=1}^{J}\left(\mathbf{Z}_j - n\bar{\mathbf{z}}_j'\right) \\ \sum_{j=1}^{J}\left(\mathbf{Z}_j' - n\bar{\mathbf{z}}_j\right) & (N-J)\mathbf{S}_{PW} \end{bmatrix}
\end{aligned} \tag{C.12}$$

We have shown that that $\mathbf{Q}_W$ from the maximally general GLS estimator can be expressed in terms of the within-cluster covariance matrix of predictors.