# Correcting Fixed Effect Standard Errors When a Crossed Random Effect was Ignored for Balanced and Unbalanced Designs

Mark H. C. Lai

University of Southern California

## Author Note

*This manuscript has been accepted for publication in the Journal of Educational and Behavioral Statistics on 03/17/2019.*

Mark H. C. Lai, Department of Psychology, University of Southern California.

Correspondence concerning this article should be addressed to Mark Lai.

Mark Lai is an Assistant Professor at the Department of Psychology, University of Southern California, 3620 South McClintock Ave., Los Angeles, LA 90089-1061; email: hokchiolai@usc.edu. His research interests are effect size and analytic issues in multilevel modeling and evaluation of measurement invariance using structural equation modeling.

# Abstract

Previous studies have detailed the consequence of ignoring a level of clustering in multilevel models with straightly hierarchical structures and have proposed methods to adjust for the fixed effect standard errors. However, in behavioral and social science research, there are usually two or more crossed clustering levels, such as when students are cross-classified by schools and neighborhoods, yet it is not uncommon that researchers focus only on one level of clustering. Using the generalized least squares framework, in this study we derive the bias in the fixed effect standard error estimators when one crossed random effect is omitted. We then showed, using data from the Scotland Neighborhood Study, how one can correct for the standard errors and obtain corrected statistical inference when a misspecified two-level model was used in a primary study, which is useful when evaluating observational studies or cluster randomized trials that ignored a crossed random effects or when conducting meta-analyses. In addition, our analytic results provide theoretical insights on how one can quantify imbalance with cross-classified data by the strength of association between the two-crossed random effects in a contingency table, and how the degree of imbalance relates to the correction factor for the fixed effect standard errors.

*Keywords:* cross-classified, multilevel model, correction, standard error, misspecification

# Correcting Fixed Effect Standard Errors When a Crossed Random Effect was Ignored for Balanced and Unbalanced Designs

Data collected in behavioral and social science research commonly are clustered into multiple levels, such as schools, classrooms, or neighborhoods, and are thus dependent. It is well known that when the dependencies among the observations are not correctly taken into account, the standard errors (*SE*) of the regression coefficients will usually be underestimated (e.g., Hox, Moerbeek, & Van de Schoot, 2018), and the estimator of the coefficients is generally less efficient (i.e., with larger sampling variability, Maas & Hox, 2005) than using approaches designed to handle clustered data, such as multilevel modeling. Previous research has studied the impact of ignoring clustering in various multilevel designs, including two-level designs (Berkhof & Kampen, 2004; Snijders, 2005), three- and four-level designs (Moerbeek, 2004; Van den Noortgate, Opdenakker, & Onghena, 2005), and cross-classified designs (Luo & Kwok, 2009). Such research both contributed to methodological understanding of multilevel models and addressed the practical need of obtaining corrected *SE* and statistical inferences when raw data are not available for a primary study that did not account for all levels of clustering.

In this article, we analytically examined the impact of ignoring one of two crossed random effects of clustering for designs with unbalanced cell sizes. Specifically, we derived a close form expression of the bias in the estimated *SE* and sampling variance (i.e., $SE^2$) of the fixed effects, and demonstrated how one can obtain corrected *SE*s and statistical inference as was done in Hedges (2007a) for two-level designs. We also showed that unbalanced cell sizes induced an association on the two crossed random effects on the outcome, and the impact of ignoring one level of clustering is generally smaller.

This article is relevant both theoretically and practically. First, from a methodological stand point, it provides a deeper understanding on cross-classified multilevel designs by specifically showing the relative contribution of two crossed random effects on the fixed effect *SE*s of predictors at different levels. Second, it shows that imbalance of a cross-classified design, expressed in terms of unequal cell sizes, can be measured by indices developed for contingency tables, and that the degree of imbalance directly factors in the bias of *SE*s when one crossed random effect is ignored. Third, on a more practical stand point, it illustrates how fixed effect *SE*s can be corrected when data are analyzed with a crossed random effect ignored in a primary study, which is not uncommon (e.g., school-randomized trials with neighborhoods ignored). Although ideally one can reanalyze the original data using CCREMs, in some situations there is a need to evaluate the original study, compare it to another study, or synthesize it with other studies (e.g., in meta-analyses) when the raw data are not available (Hedges, 2007a). In those situations the corrected *SE*s discussed in this article will be particularly useful.

## Cross-Classified Random Effect Models (CCREMs)

With the increased accessibility of software packages with multilevel modeling capability (e.g., `SPSS MIXED`, Norusis, 2012; `SAS PROC MIXED`, Littell, 2006; `HLM`, Raudenbush, Bryk, Cheong, Congdon, & Du Toit, 2011; `MLWiN`, Rasbash, Steele, Browne, & Goldstein, 2017; `nlme`, Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2018; and `lme4`, Bates, Mächler, Bolker, & Walker, 2015 in R), for the past decade applied educational researchers have become more aware of the need to model the effect of clustering such as schools or classrooms. On the other hand, educational data can have a highly complex structure, as student-level variables can be affected by multiple contextual factors such as classrooms, schools, neighborhoods, and so forth. Whereas research methodologists have proposed CCREMs (Goldstein, 1994; Raudenbush, 1993) to handle multiple random effects that are *crossed* and do not follow a hierarchical structure (e.g., neighborhoods and schools where students from the same neighborhood attend multiple schools, and students in the same school come from multiple neighborhoods), not until recently did these CCREMs become more popular. The increasing popularity of CCREM is evident as a simple search in the Educational Research Information Center (ERIC) database with the keyword “cross-classified” and “crossed random” found only 11 journal articles in the period 1993–2005 but 72 journal articles in the period 2006–2016.

The increasing adoption of CCREMs means that (a) it is of paramount importance to understand the statistical properties of such models, and (b) there is a need to obtain corrected statistical inference for older literature when CCREMs were not yet developed and routinely used. In a Monte Carlo simulation study, Meyers and Beretvas (2006) found that when a crossed random effect was ignored, the *SE*s of fixed effect estimates for predictors at the ignored level were generally underestimated, and the magnitude of *SE* bias increased with a higher intraclass correlation and a larger cluster size for the ignored level, and decreased when the residuals of the two crossed random effects were correlated. Luo and Kwok (2009) provided analytic results for the impact of ignoring a crossed random effect in three-level CCREMs with balanced designs, and demonstrated the downward biases in the *SE*s with unbalanced designs using Monte Carlo simulation. Luo and Kwok (2012) further expanded the results to longitudinal settings where students are cross-classified by multiple school memberships, and showed with simulation studies that the variance components were incorrectly estimated and the fixed effect *SE*s were underestimated when students' mobility was not correctly modeled by ignoring the cross-classification of the clustering levels.

We aimed to achieve several purposes in this paper. First, we derived a general form of the bias on the *SE*s of the fixed effects coefficients when a two-level hierarchical linear model (HLM), which only accounts for one level of clustering, is used to model data cross-classified by two levels, and showed how such bias can be corrected. The closed form expression of the bias was obtained using generalized least squares (GLS), which is asymptotically equivalent to the maximum likelihood (ML) and restricted maximum likelihood (REML) estimators, and they are identical in finite samples when the variance components are known (Snijders & Bosker, 1993). Then we made explicit the link between unbalanced cross-classified structure (i.e., with unbalanced cell sizes) and the association for the contingency table of the two crossed random effects, and used that to express the corrections for predictors defined at different levels of clustering. Finally, we demonstrated how the fixed effect *SE*s can be corrected using an empirical example with the Scotland Neighborhood Study data (Garner & Raudenbush, 1991; Raudenbush & Bryk, 2002).

## Model and Notation

Consider a multilevel model with two crossed random effects, *A* and *B*. For simplicity, we assume that the model contains only one single fixed predictor, *x*, with a fixed effect parameter $\gamma$, but *x* can be defined at any level. The results should generalize to multiple predictors as long as the multicollinearity among them are taken into account in the misspecified model. Let *a* be the number of clusters for level *A* and *b* be the number of clusters for level *B* in the sample. A general CCREM model can be formulated as

$$\mathbf{y} = \mathbf{x}\gamma + \mathbf{Z}_1\mathbf{u} + \mathbf{Z}_2\mathbf{v} + \boldsymbol{\varepsilon}, \tag{1}$$

where **y** and **x** are an $N \times 1$ column vectors of, respectively, the scores of the response variables and of the fixed predictor, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are the $N \times a$ and $N \times b$ matrices of 0s and 1s denoting the group membership for each of the *N* observations on level *A* and level *B*, and **u** and **v**, are column vectors of length *a* and *b* of random effects for *A* and *B*, and $\boldsymbol{\varepsilon}$ is the column vector of length *N* of the within-cell level error terms. The random effects and errors are assumed independent and normally distributed such that $\mathbf{u} \sim N(0, \tau_A^2\mathbf{I}_a)$, $\mathbf{v} \sim N(0, \tau_B^2\mathbf{I}_b)$, and $\boldsymbol{\varepsilon} \sim N(0, \sigma^2\mathbf{I}_N)$. Note that given the complexity of CCREM, we limit our discussion to models with no random slopes. In addition, the model assumes that the interaction between *A* and *B* is zero, as a closed form expression for the bias when the interaction is nonzero generally does not exist with ML/REML estimation. However, Shi, Leite, and Algina (2010) and Wallace (2015) found that in CCREM with two crossed random effects, ignoring a non-zero interaction between the crossed random effects only biased the random effect variance estimates, but not the fixed effect point and *SE* estimates. This implies that the presence of an interaction between two crossed random effects had very little impact on the bias expressions and corrections provided below, so the correction procedure we proposed in this paper should still give reasonable results when both a crossed random effect and its interaction with the other crossed random effect was omitted.

Under the CCREM in equation (1), the sampling variance of **y** is $\mathrm{Var}(\mathbf{y}) = \mathbf{V} =$

$\sigma^2\mathbf{I} + \tau_A^2\mathbf{Z}_1'\mathbf{Z}_1 + \tau_B^2\mathbf{Z}_2'\mathbf{Z}_2$. Under the assumption that the matrix $\mathbf{V}$ is positive definite, the GLS estimator of $\gamma$, denoted as $\hat{\gamma}$, is $\left(\mathbf{x}'\mathbf{V}^{-1}\mathbf{x}\right)^{-1}\mathbf{x}'\mathbf{V}^{-1}\mathbf{y}$, with sampling variance $\mathrm{Var}(\hat{\gamma}) = \left(\mathbf{x}'\mathbf{V}^{-1}\mathbf{x}\right)^{-1}$ (Snijders & Bosker, 1993). With ML/REML, the estimated variance of $\hat{\gamma}$ can be obtained by substituting the ML/REML estimates of the variance components, that is,

$$\hat{\mathrm{Var}}(\hat{\gamma}) = \left(\mathbf{x}'\hat{\mathbf{V}}^{-1}\mathbf{x}\right)^{-1},$$

where $\hat{\mathbf{V}} = \hat{\sigma}^2\mathbf{I} + \hat{\tau}_A^2\mathbf{Z}_1'\mathbf{Z}_1 + \hat{\tau}_B^2\mathbf{Z}_2'\mathbf{Z}_2$. The *SE* estimate of $\hat{\gamma}$ is then the square root of $\hat{\mathrm{Var}}(\hat{\gamma})$.

## Relative Bias and Correction of Fixed Effect Standard Errors When Ignoring a Crossed Random Effect

When the model is misspecified as a two-level HLM with one of the crossed random effect omitted, the estimators of variance components and fixed effects change. In the discussion below we assumed that the variance components in the correctly specified model are known so that we can expression the *SE* bias as a function of them. Without loss of generality, we assume that level *B* is omitted. Let $\tilde{\tau}_A^2$ and $\tilde{\sigma}^2$ be the variance component estimators of level *A* and the within-cell level, and $\tilde{\gamma}$ be the fixed effect estimator, under the misspecified model. The misspecified model implies that $\mathrm{Var}(\mathbf{y}) = \tilde{\mathbf{V}} = \tilde{\sigma}^2\mathbf{I} + \tilde{\tau}_A^2\mathbf{Z}_1'\mathbf{Z}_1$, and the GLS estimator for the fixed effect is $\tilde{\gamma} = \left(\mathbf{x}'\tilde{\mathbf{V}}^{-1}\mathbf{x}\right)^{-1}\mathbf{x}'\tilde{\mathbf{V}}^{-1}\mathbf{y}$. Note that $\tilde{\gamma}$ is still consistent for $\gamma$, as also demonstrated in Luo and Kwok (2009) and Meyers and Beretvas (2006). The *true* sampling variance of the $\tilde{\gamma}$, however, has a sandwich form:

$$\mathrm{Var}(\tilde{\gamma}) = (\mathbf{x}'\tilde{\mathbf{V}}^{-1}\mathbf{x})^{-1}\mathbf{x}'\tilde{\mathbf{V}}^{-1}\mathbf{V}\tilde{\mathbf{V}}^{-1}\mathbf{x}(\mathbf{x}'\tilde{\mathbf{V}}^{-1}\mathbf{x})^{-1}. \tag{2}$$

On the other hand, under the misspecified model, the *estimated* variance of $\tilde{\gamma}$, denoted as $\hat{\mathrm{Var}}(\tilde{\gamma})$, is generally biased and an underestimate, with

$$\hat{\mathrm{Var}}(\tilde{\gamma}) = \left(\mathbf{x}'\tilde{\mathbf{V}}^{-1}\mathbf{x}\right)^{-1}, \tag{3}$$

Thus, comparing (2) and (3), and define

$\mathbf{V}^O = \mathbf{V} - \tilde{\mathbf{V}} = (\sigma^2 - \tilde{\sigma}^2)\mathbf{I} + (\tau_A^2 - \tilde{\tau}_A^2)\mathbf{Z}_1\mathbf{Z}_1' + \tau_B^2\mathbf{Z}_2\mathbf{Z}_2'$, the bias of the estimated variance on a misspecified model is:

$$\text{Bias}[\hat{\text{Var}}(\tilde{\gamma})] = -\left(\mathbf{x}'\tilde{\mathbf{V}}^{-1}\mathbf{x}\right)^{-1}\mathbf{x}'\tilde{\mathbf{V}}^{-1}\mathbf{V}^O\tilde{\mathbf{V}}^{-1}\mathbf{x}\left(\mathbf{x}'\tilde{\mathbf{V}}^{-1}\mathbf{x}\right)^{-1}. \tag{4}$$

Because bias is not scale free, it is more instructive to look at the relative bias of the estimated fixed effect sampling variance: $\text{RBias}[\hat{\text{Var}}(\tilde{\gamma})] = \text{Bias}[\hat{\text{Var}}(\tilde{\gamma})]/\text{Var}(\tilde{\gamma})$. Using (4), the relative bias expression can be simplified as

$$\text{RBias}[\hat{\text{Var}}(\tilde{\gamma})] = -\frac{\mathbf{x}'\tilde{\mathbf{V}}^{-1}\mathbf{V}^O\tilde{\mathbf{V}}^{-1}\mathbf{x}}{\mathbf{x}'\tilde{\mathbf{V}}^{-1}\mathbf{V}\tilde{\mathbf{V}}^{-1}\mathbf{x}}. \tag{5}$$

When variance components are estimated instead of known, we can substitute the variance components with the ML or REML estimates to approximate the relative bias, as long as the variance component estimators are consistent. Note that close-form expression of (5), which involves inverse of $\tilde{\mathbf{V}}$, generally does not exist except for some special cases, including when the cluster sizes are equal for each crossed random effect, which we will discuss in the next section. Once the relative bias is estimated, one can correct $\hat{\text{Var}}(\tilde{\gamma})$ as

$$\text{Corrected}\ \hat{\text{Var}}(\tilde{\gamma}) = \frac{\hat{\text{Var}}(\tilde{\gamma})}{1 + \text{RBias}[\hat{\text{Var}}(\tilde{\gamma})]}, \tag{6}$$

and taking the square root gives the corrected *SE* of $\tilde{\gamma}$.

Although expressions (5) and (6) are solvable when one has the raw data, they do not provide insights into how cluster sizes, degree of imbalance, and magnitudes of variance components contribute to the degree of bias on the fixed effects. To do so, we need to first express the estimated variance components under the misspecified HLM in terms of the parameters under CCREM.

## Estimated Variance Components Under a Misspecified Model

Let $n_{ij}$ be the cell size for the intersection of the $i$th cluster at level $A$ and the $j$th cluster at level $B$, $n_{i.}$ be the size of the $i$th cluster at level $A$, and $n_{.j}$ be the size of the $j$th cluster at level $B$. For a CCREM with equal cluster sizes for both levels $A$ and $B$ such that $n_{i.} = c_A$ and $n_{.j} = c_B$ but with cell sizes not necessarily constant, if we ignore level $B$, one ends up with a regular two-level HLM with equal cluster sizes, where close-form expression of $\hat{\mathrm{Var}}(\tilde{\gamma})$ exists given the $\tilde{\sigma}^2$ and $\tilde{\tau}_A^2$ estimates. Define $\theta_A = \tau_A^2/\sigma^2$, $\theta_B = \tau_B^2/\sigma^2$, and $\kappa_A = c_A\theta_A$ and $\kappa_B = c_B\theta_B$. As detailed in Appendix A, under the misspecified model the variance component estimates become

$$\tilde{\tau}_A^2 = \frac{\sigma^2}{c_A}\{\kappa_A + [\varphi^2 s_1 - (1 - \varphi'^2)s_2]\kappa_B\}, \tag{7}$$

$$\tilde{\sigma}^2 = \sigma^2[1 + (1 - \varphi'^2)s_2\kappa_B], \tag{8}$$

where $s_1 = 1/(a - 1)$ and $s_2 = (b - 1)/(N - a)$ are known constants, and

$$\varphi^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{n_{ij}^2}{c_A c_B} - 1, \tag{9}$$

$$\varphi'^2 = \varphi^2/(b - 1), \tag{10}$$

which are two quantities that increase with larger degree of imbalance in the cell sizes.

## Quantifying the Degree of Imbalance in CCREM

Before substituting (7) and (8) into (5) to obtain the relative bias expressions, we should first make sense of what $\varphi^2$ and $\varphi'^2$ are, as it leads to insight on a way to quantify the degree of imbalance in CCREM. As to be shown below and hinted by Luo and Kwok (2009), the degree of imbalance can affect the bias on the fixed effects, yet it has otherwise not been discussed in multilevel literature.

Consider again the model defined in (1). Although $\mathbf{u}$ and $\mathbf{v}$ are assumed independent, their impact on $\mathbf{y}$ depends on $\mathbf{Z}_1$ and $\mathbf{Z}_2$ (see Figure 1). If $\mathbf{Z}_1'\mathbf{Z}_2 = n\mathbf{1}_a\mathbf{1}_b'$, meaning that each cell in the contingency table $\mathbf{Z}_1'\mathbf{Z}_2$ has the same size $n$, the design is balanced, and $\mathbf{Z}_1\mathbf{u}$ and $\mathbf{Z}_2\mathbf{v}$ are independent. Otherwise $\mathbf{Z}_1\mathbf{u}$ and $\mathbf{Z}_2\mathbf{v}$ are associated. Because $A$ and $B$ are categorical variables, one can compute the Pearson's $\chi^2$ for contingency table (Pearson, 1900) as:

$$\chi^2 = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{(n_{ij} - n_{i.}n_{.j}/N)^2}{n_{i.}n_{.j}/N}. \tag{11}$$

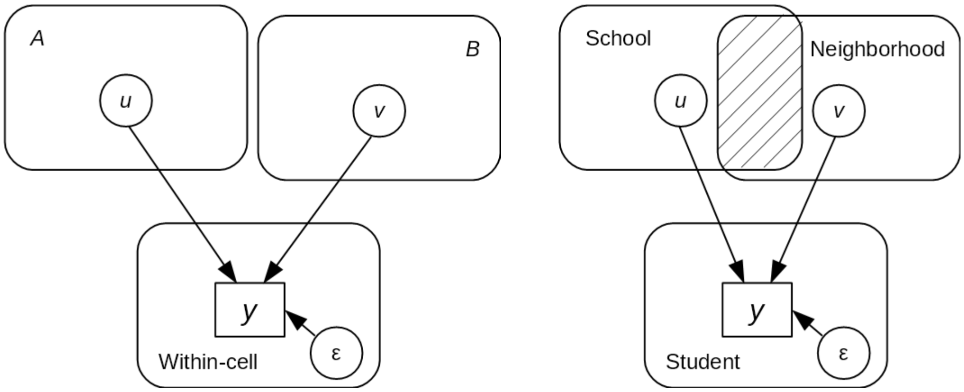
It can be shown that, when the cluster sizes are equal such that $n_{i.} = c_A$ and $n_{.j} = c_B$,

$$\chi^2 = N\varphi^2 \tag{12}$$

where $\varphi^2$ is defined in equation (9). Equations (7) to (12) show that $\tilde{\tau}_A^2$ and $\tilde{\sigma}^2$ are direct functions of the strength of association in the contingency table of $A$ and $B$. Furthermore, if we look at the definition of Cramér's $V$ (denoted as $V_c$ in this article to distinguish it from $\mathbf{V} = \mathrm{Var}[\mathbf{y}]$; cf. Grissom & Kim, 2012)—an effect size index for contingency table on a scale of 0 to 1 with 0 meaning balanced and 1 meaning maximum imbalance, we have

$$\begin{aligned} V_c^2 &= \frac{\chi^2/N}{\min(a-1, b-1)} = \frac{\varphi^2}{\min(a-1, b-1)} \\ &= \begin{cases} \varphi^2 s_1 & \text{if } a \le b \\ \varphi'^2 & \text{if } a > b \end{cases}. \end{aligned}$$

As $0 \le V_c \le 1$ and $(a-1)/(b-1) > 1$ when $a > b$, we also have $0 \le \varphi^2 s_1 \le 1$ and $0 \le \varphi'^2 \le 1$. This implies that one can directly use $V_c$ to quantify the degree of imbalance for cross-classified data. Three special cases are worth mentioning.



(a) (b)

*Figure 1.* Cross-classified random effect models of outcome *y* with (a) a balanced design (i.e., fully cross-classified) and (b) an unbalanced design (i.e., partially cross-classified). Boxes with sharp corners denote observed variables, circles denote random effects and error term, and boxes with round corners are used to separate different levels of clustering. Note that the two crossed random effects, *u* and *v*, are assumed independent.

**$V_c = 0$.** When $V_c = 0$, $\varphi^2 = \varphi'^2 = 0$, and the data are balanced or fully cross-classified (see Figure 2a for an example).

**$V_c = 1, a < b$.** When $V_c = 1$, $a < b$, we have $\varphi^2 = a - 1$. From equation (9) this implies that $\sum_{i=1}^{a} \sum_{i=1}^{b} n_{ij}^2 = c_B N$, and the only way this happens is that $B$ is nested within $A$ (see Figure 2c for an example). Therefore, in this situation, omitting level $B$ is equivalent to omitting the *middle* level in a three-level hierarchical model. Moerbeek (2004) and Van den Noortgate et al. (2005) showed that ignoring a middle level generally results in underestimated fixed effect *SE*s at that level, but may also affect fixed effect *SE*s at the top and the lowest levels.

**$V_c = 1, a > b$.** Similarly, when $V_c = 1$, $a > b$, we have the situation where $A$ is nested within $B$ (see Figure 2d for an example). Therefore, omitting level $B$ is equivalent to omitting the *top* level in a three-level hierarchical model. Moerbeek (2004) and Van den Noortgate et al. (2005) showed that ignoring a top level generally results in underestimated fixed effect *SE*s at that level, but may also affect fixed effect *SE*s at the middle level.

|  | B1 | B2 | B3 |
|---|---|---|---|
| A1 | 3 | 3 | 3 |
| A2 | 3 | 3 | 3 |
| A3 | 3 | 3 | 3 |
| A4 | 3 | 3 | 3 |
| A5 | 3 | 3 | 3 |
| A6 | 3 | 3 | 3 |

(a)

|  | B1 | B2 | B3 |
|---|---|---|---|
| A1 | 8 | 1 | 0 |
| A2 | 7 | 1 | 1 |
| A3 | 0 | 6 | 3 |
| A4 | 1 | 6 | 2 |
| A5 | 0 | 2 | 7 |
| A6 | 0 | 0 | 9 |

(b)

|  | B1 | B2 | B3 | B4 | B5 | B6 |
|---|---|---|---|---|---|---|
| A1 | 9 | 9 | 0 | 0 | 0 | 0 |
| A2 | 0 | 0 | 9 | 9 | 0 | 0 |
| A3 | 0 | 0 | 0 | 0 | 9 | 9 |

(c)

|  | B1 | B2 | B3 |
|---|---|---|---|
| A1 | 9 | 0 | 0 |
| A2 | 9 | 0 | 0 |
| A3 | 0 | 9 | 0 |
| A4 | 0 | 9 | 0 |
| A5 | 0 | 0 | 9 |
| A6 | 0 | 0 | 9 |

(d)

*Figure 2*. Contingency tables of two crossed random effects $A$ and $B$ for (a) fully cross-classified data, Cramér's $V$ ($V_c = 0$), (b) partially cross-classified data, $V_c = 0.728$, (c) three-level data, $V_c = 1$ with $B$ nested within $A$, and (d) three-level data, $V_c = 1$ with $A$ nested within $B$.

Finally, if $V_c = 1$ and $a = b$, that $A$ and $B$ are identical and so one of them is redundant.

Whereas a balanced design results in $V_c = 0$, in real observational data sets with crossed random effects of clustering, the degree of imbalance is usually quite high. For example, for the Scotland Neighborhood data used in the demonstration later in the current paper, $V_c = .875$ for the contingency table between neighborhood and schools. For the 1988 National Educational Longitudinal Study (Ingles, Abraham, Karr, Spencer, & Frankel, 1990) data set, $V_c = .952$ for the contingency table between middle school and high school.

## Relative *SE* Bias and Corrected *SE* for Predictors at Different Levels

In this section we present simplified relative *SE* bias expression for predictors at different levels, based on equation (5). To precisely distinguish the predictors at different levels, and drawing on the notations and concepts introduced by Berkhof and Kampen (2004), we define $SS_x$ as the sum of squares of the predictor $X$ and $f_{xA} = SS_{xA}/SS_x$ as the proportion of sum of squares of the predictor $X$ at level $A$. Note that $f_{xA}$ is sample specific and takes value between 0 and 1. If $X$ is assumed stochastic and the intercept is included in the model, $f_{xA}$ is the proportion of variance at level $A$ for $X$. Similarly, define $f_{xB} = SS_{xB}/SS_x$. Then, in a design with two crossed random effects, a predictor $X$ can be categorized into four types:

1. a grand mean or an intercept, which corresponds to a column vector of 1s,
2. a level-$A$ predictor with $f_{xA} = 1$ and $0 \leq f_{xB} < 1$,
3. a level-$B$ predictor with $f_{xB} = 1$ and $0 \leq f_{xA} < 1$,

4. a level-1 predictor with $0 \leq f_{xA} < 1$ and $0 \leq f_{xB} < 1$.

Note that when the cell sizes are equal, for a level-$A$ predictor, $f_{xB} = 0$; otherwise $f_{xB} > 0$ even for a level-$A$ predictor, and it is closer to 1 with increasing $V_c$. Similarly, $f_{xA} > 0$ for a level-$B$ predictor unless the cell sizes are equal.

From the matrix expression in (5), with some involved substitutions shown in Appendix B, the relative bias expression for predictors in closed form with any values of $f_{xA}$ and $f_{xB}$ is

$$\text{RBias}[\hat{\text{Var}}(\tilde{\gamma})] = \frac{\kappa_B[(1-\varphi'^2)s_2(1+\tilde{\kappa}_A)^2(1-f_{xA}) + \varphi^2 s_1 f_{xA} - (1+\tilde{\kappa}_A)^2 f^O_{xB}]}{(1+\tilde{\kappa}_A)^2(1-f_{xA}+\kappa_B f^O_{xB}) + (1+\kappa_A)f_{xA}}, \tag{13}$$

where $\tilde{\kappa}_A = c_A\tilde{\theta}_A$ and $\tilde{\theta}_A = \tilde{\tau}^2_A/\tilde{\sigma}^2$ (i.e., the analogue of $\kappa_A$, but based on the misspecified model), $f^O_{xB} = f_{xB} - 2\frac{\tilde{\kappa}_A}{1+\tilde{\kappa}_A} f_{xA} f_{xB} + \frac{\tilde{\kappa}^2_A}{(1+\tilde{\kappa}_A)^2}\mathbf{x}'_A\mathbf{Z}_2\mathbf{Z}'_2\mathbf{x}_A/c_B/SS_x$ , $\mathbf{x}_A = \mathbf{Z}_1\mathbf{Z}'_1\mathbf{x}/c_A$ a $N \times 1$ column vector containing the cluster mean at level $A$ of each observation. When $f_{xA} = 1$, $f^O_{xB}$ is reduced to $f_{xB}$; otherwise $f^O_{xB} > f_{xB}$, which brings the relative bias closer to zero. Therefore, the relative bias on $\hat{\text{Var}}(\tilde{\gamma})$ is mainly a function of the clustering effects of $A$ and $B$ (i.e., $\tilde{\kappa}_A$ and $\kappa_B$), the degree of imbalance in the cell sizes ($\varphi^2$ and $\varphi'^2$), and the proportion of sum of squares of the predictor at different levels ($f_{xA}$ and $f_{xB}$). After the relative bias is obtained, the fixed effect sampling variance and the *SE* can be corrected using (6). Below we examine the relative sampling variance bias of the fixed effect estimators for each type of predictor for both balanced and unbalanced designs.

**Grand mean/intercept.** First, for the special case of the grand mean in an unconditional model or the intercept in a conditional model, which corresponds to a column vector of 1s in the design matrix **x**, we have $f_{xA} = f_{xB} = 1$ such that all level of clustering contributed to the sampling variance. The relative bias expression in (13) then simplifies to

$$\text{RBias}[\hat{\text{Var}}(\tilde{\gamma})] = \frac{[\varphi^2 s_1 - 1]\kappa_B}{1+\kappa_A+\kappa_B}. \tag{14}$$

Therefore, the estimated sampling variance of the grand mean or the intercept is negatively biased with relative bias equals $-\kappa_B/(1+\kappa_A+\kappa_B)$ for a balanced design with $\varphi^2 = 0$, and is unbiased only when $\varphi^2 s_1 = 1$, which only happens when $B$ is completely nested in $A$ (i.e., the omitted level is the middle level in a three-level data structure). Asymptotically, from equation (14), $\lim_{a\to\infty} \text{RBias}[\hat{\text{Var}}(\tilde{\gamma})] \to -1$, and $\lim_{b\to\infty} \text{RBias}[\hat{\text{Var}}(\tilde{\gamma})] \to 0$, so, for fixed cell sizes, the bias is generally negligible when number of clusters is large for the omitted level (i.e., *B*). Also, the relative bias does not depend on the degree of unbalanced cell size sizes when *a* or *b* is large.

**Level-*A* predictors.** For variables defined at level *A* (i.e., $f_{xA} = 1$), the relative bias expression on the variance of the fixed effect estimators simplifies to

$$\text{RBias}[\hat{\text{Var}}(\tilde{\gamma}^{(A)})] = \frac{[\varphi^2 s_1 - f_{xB}]\kappa_B}{1 + \kappa_A + f_{xB}\kappa_B}, \tag{15}$$

which is similar to that for grand mean/intercept. Although both Luo and Kwok (2009) and Meyers and Beretvas (2006) both suggested that when clustering of *A* was accounted for, the *SE* bias for variables at the level *A* was negligible, from the above analytical expression one can see that such a finding in the previous literature was only true when $f_{xB}$ and $\varphi^2 s_1$ are similar or when $\kappa_B$ is small relative to $\kappa_A$ (so that the numerator is small). One special case is when the design is balanced with $\varphi^2 = 0$, which by definition implies that $f_{xB} = 0$, and so the relative bias is zero. Also, if *X* is a stochastic level-*A* variable, $f_{xB}$ converges to $\varphi^2 s_1$ in large sample, so the relative bias will be close to zero (as were the case in previous simulation studies). When *X* is fixed and the design is unbalanced, however, $\varphi^2 s_1$ can be smaller or larger than $f_{xB}$. Figure 3 illustrated equation (15) across different combinations of *a*, *b*, average cell size ($\bar{n}$), degree of unbalanced cluster sizes, and conditions of $f_{xB} - \varphi^2 s_1$, with $\theta_A = 2/7$ and $\theta_B = 1/7$ (which corresponds to conditional intraclass correlations of .20 and .10 for *A* and *B*, respectively). The figure shows that, generally, the relative bias of the sampling variance of the fixed effects at level *A* is relatively small when level *B* is omitted. Asymptotically, $\lim_{a\to\infty} \text{RBias}[\hat{\text{Var}}(\tilde{\gamma})] \to -1$ and $\lim_{b\to\infty} \text{RBias}[\hat{\text{Var}}(\tilde{\gamma})] \to 0$, so the bias does not depend on the degree of imbalance of cell sizes when *a* or *b* is large, as in the case for the grand mean/intercept.

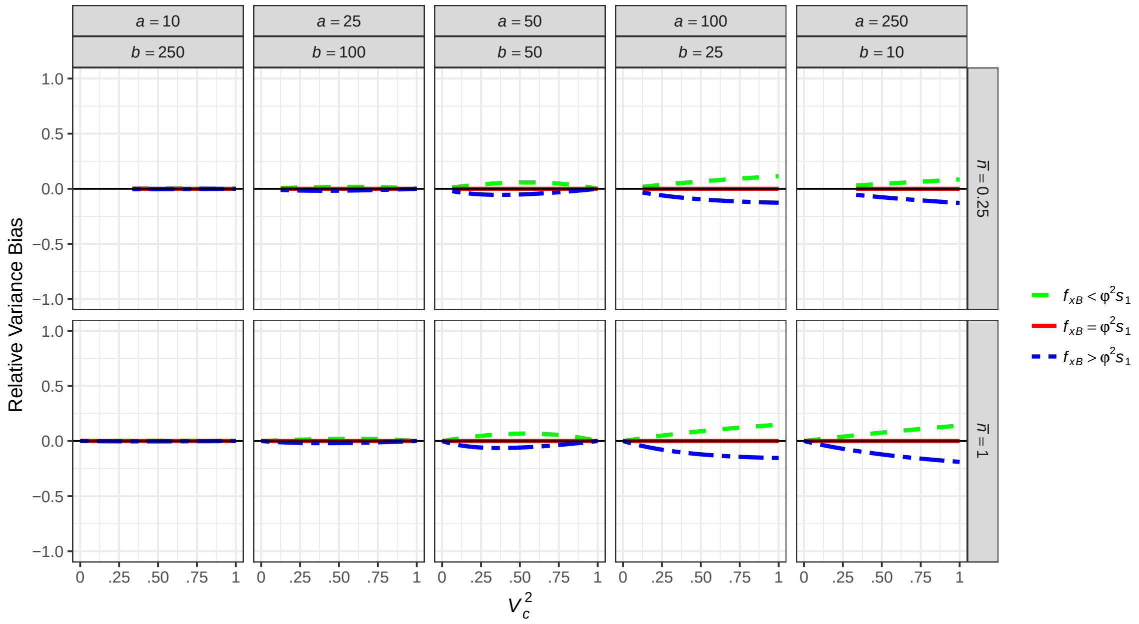


*Figure 3*. Relative sampling variance bias of fixed effect estimators for a predictor at level $A$ when ignoring a crossed random effect $B$. $a$, $b$ = number of clusters at level $A$ and level $B$. $s_1 = 1/(a-1)$. $V_c^2$ = squared Cramér's $V$. $\bar{n}$ = average cell size. $f_{xB}$ = proportion of sum of squares of the predictor at level $B$. $\varphi^2$ is defined in equation (9).

**Level-*B* predictors.** For variables defined at level $B$ (i.e., $f_{xB} = 1$), the expression of the relative variance bias of the fixed effect can be obtained by substituting $f_{xB} = 1$ into expression (13), with $f_{xB}^O = 1 - 2\frac{\tilde{\kappa}_A}{1+\tilde{\kappa}_A} f_{xA} + \frac{\tilde{\kappa}_A^2}{1+\tilde{\kappa}_A}\mathbf{x}_A'\mathbf{Z}_2\mathbf{Z}_2'\mathbf{x}_A/c_B/SS_x$, which is generally smaller for larger $f_{xA}$. For a balanced design, $\varphi^2 = \varphi'^2 = f_{xA} = 0$ and $f_{xB}^O = 1$, so the relative bias expression simplifies to

$$\text{RBias}[\hat{\text{Var}}(\tilde{\gamma}^{(B)})] = \frac{\kappa_B(s_2 - 1)}{1 + \kappa_B},$$

which increases with larger $a$ but decreases with larger $b$. For unbalanced designs, the $\text{RBias}[\hat{\text{Var}}(\tilde{\gamma})]$ expression does not simplify much. Therefore, it is instructive to plot the relative variance bias for different degrees of imbalance and cluster size combinations, as shown in Figure 4. We approximated $f_{xB}^O$ by $\left(1 - \frac{\tilde{\kappa}_A f_{xA}}{1+\tilde{\kappa}_A}\right)^2$, which holds in large sample for a stochastic level-$B$ predictor. As the value of $f_{xA}$ is close to $\varphi^2 s_1$ for a stochastic level-$B$ predictor, in the figure we showed relative biases across three conditions: $f_{xA} > \varphi^2 s_1$, $f_{xA} = \varphi^2 s_1$, and $f_{xA} < \varphi^2 s_1$. The relative bias generally is larger for larger cluster size at the omitted level (i.e., increases in $a$ or

$\bar{n}$, $c_B = a\bar{n}$), and under such conditions the degree of imbalance exerts only negligible influence on the relative bias of the sampling variance. When $a \leq b$, increases in the imbalance of cell sizes resulted in less biased sampling variance estimates, as some of the missing variance in $B$ is captured by $A$ due to their increased association. When $\varphi^2 s_1 = 1$ such that $B$ is nested within $A$, meaning that $B$ is the middle level, ignoring $B$ does not affect the *SE* estimate of fixed effects at level $B$, as long as $A$, the top level, was included.

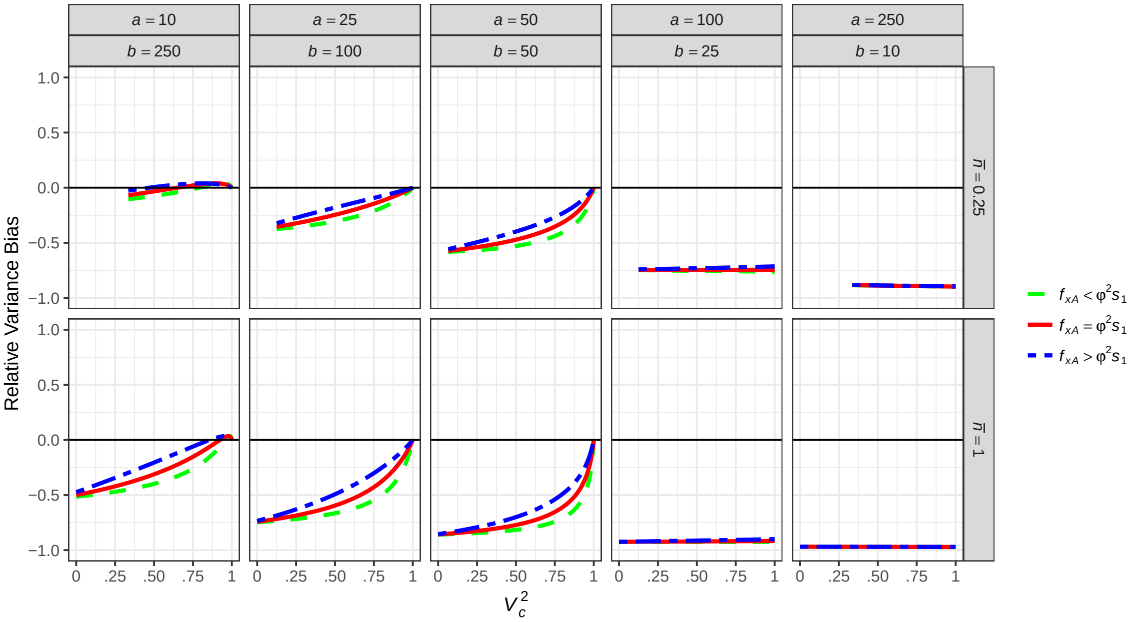


*Figure 4.* Relative sampling variance bias of fixed effect estimators for a predictor at level $B$ when ignoring a crossed random effect $B$. $a$, $b$ = number of clusters at level $A$ and level $B$. $s_1 = 1/(a-1)$. $V_c^2$ = squared Cramér's $V$. $\bar{n}$ = average cell size. $f_{xB}$ = proportion of sum of squares of the predictor at level $A$. $\varphi^2$ is defined in equation (9).

**Level-1 predictors.** For general level-1 predictors, $f_{xA} > 0$ and $f_{xB} > 0$. In other words, similar to the outcome variable $Y$, a level-1 predictor $X$ can have varying degree of intraclass correlations at level $A$ and at level $B$. If $X$ is a stochastic variable with no level-$A$ and level-$B$ variance components at the population level, $f_{xB}$ and $f_{xA}$ both will converge to 0 as sample size increases, so will $f^O_{xB}$; from (13) it can then be shown that the relative bias will converge to 0, and correction will not be needed.[1] Otherwise the bias needs to be computed based on the general form in (13). As $f_{xB}$ gets closer to 1, the variable becomes closer to a level-$B$ predictor so that the downward bias on the fixed effect sampling variance is larger, similar to the pattern shown in Figure 4.

## Bias With Unequal Cluster Sizes

Similar to previous analytic studies on fixed effect estimators in the presence of clustering (e.g., Berkhof & Kampen, 2004), we assume that the cluster sizes for each crossed random effect were equal (although the cell sizes are not equal unless $V_c = 0$) when deriving expression (13) and the special cases, as otherwise the standard error estimates of the ML/REML fixed effect estimators involve intractable matrix algebra. However, given that cluster sizes are usually not equal in real data, it is important to ask how the above discussion holds with unequal cluster sizes. To shed some lights on that question, we simulated data from a CCREM model with $N$ = 5,600, $a = 200$, $b = 10$, and unbalanced cluster sizes for levels $A$ and $B$ with largest and smallest cluster sizes of 213 and 19 (∼ a 11 to 1 ratio), and 1486 and 377 (∼ a 4 to 1 ratio). We set $\tau_A^2 = 0.4$, $\tau_B^2 = 0.1$, and $\sigma^2 = 1$, generated a normally distributed predictor $X$ that was either a level-1, level-$A$, or level-$B$ variable, under either a balanced ($V_c = 0$) or unbalanced ($V_c = .84$) design. With 2,000 replications, we compared the empirical *SE*, the average uncorrected *SE*, and the average corrected *SE*s with level $B$ omitted. Specifically, when $X$ was either a level-1 or a level-$A$ predictor, based on our previous discussions the estimated *SE* of $X$ is expected to have little bias when level $B$ was ignored, and the simulation confirmed this with relative biases of $-1.5\%$ to $2\%$ for the uncorrected *SE* and $-2.3\%$ to $6.1\%$ for the corrected *SE*. When $X$ was a level $B$ predictor, however, the uncorrected *SE* showed strong downward relative biases of $-85.8\%$ and $-79.6\%$ for balanced and unbalanced designs, whereas the corrected *SE* had much smaller relative biases of $2.8\%$ and $-12.2\%$. Therefore, the simulation suggested that the corrected *SE* can correct most of the bias even with with unequal cluster sizes.

# Real Data Illustration

To illustrate the application of our analytic results for estimating the biases in the fixed effect *SE*s with a crossed random effect omitted and to obtain corrected *SE*s in real data, we used data from the Scotland Neighborhood Study (Garner & Raudenbush, 1991; Raudenbush & Bryk, 2002) consisting of 2,310 student cross-classified by 524 neighborhoods (with 1 to 16 students in one neighborhood) and 17 schools (with 22 to 286 students in one school). A hypothetical research question is whether neighborhood-level social deprivation (`DEPRIVE`; $M = 0.021$, $SD = 0.622$) can predict student-level total attainment (`ATTAIN`; $M = 0.116$, $SD = 1.00$). We compared three analytical models in this illustration: a two-level HLM only modeling the clustering due to neighborhoods (HLM-N), a two-level HLM only modeling the clustering due to schools (HLM-S), and a two-level CCREM modeling both clusterings due to neighborhoods and schools (CCREM). We analyzed the three models using `lme4`, and the results were shown in Table 1. For HLM-N, the fixed effect estimate of `DEPRIVE` was $\tilde{\gamma}_{01} = -0.521$, $SE = 0.038$; for HLM-S, $\tilde{\gamma}_{01} = -0.473$, $SE = 0.033$; for CCREM, $\hat{\gamma}_{01} = -0.466$, $SE = 0.038$. Just looking on the estimated *SE*s one may be tempted to think that the model ignoring the school effect better capture the sampling variability in `DEPRIVE` as the *SE*s were similar for HLM-N and CCREM.

However, the fixed effect estimators in all three models are different with different true sampling variances, and the estimator under CCREM is likely to be the most efficient. Therefore, even though $\hat{SE}(\tilde{\gamma}_{01})$ under HLM-N was similar to $\hat{SE}(\hat{\gamma}_{01})$ under CCREM, the former underestimated the true sampling variance of $\tilde{\gamma}_{01}$, a less efficient estimator. To verify this, we treated the CCREM as the true model and generated 20,000 data sets, and obtained the empirical standard errors for the fixed effect estimators as the standard deviation of the $\gamma_{01}$ estimates across all simulated data sets for all three models; the results were shown in the "Monte Carlo *SE*" row in Table 1. It was found that the true *SE* of $\tilde{\gamma}_{01}$ under HLM-N was 0.0475, so the estimated *SE* had a relative bias of $0.038/0.0475 - 1 = -.20$. On the other hand, the true *SE* of $\tilde{\gamma}_{01}$ under HLM-S was 0.0387, so the estimated *SE* had a relative bias of $0.033/0.0387 - 1 = -.15$. Therefore, in this situation had one omitted either the school or the neighborhood level, the estimated *SE* for `DEPRIVE` would be too small, leading to confidence interval being too narrow and inflated chances of committing a Type I error, so correction is needed to obtain proper statistical inference. Also, even though `DEPRIVE` is a neighborhood-level variable, in this example the bias on the estimated

*SE* was larger when the school level was ignored than when neighborhood level was ignored.

The stronger influence of the school level on a neighborhood-level predictor can be understood by computing $\hat{\kappa}_{\text{NEIGH}}$ and $\hat{\kappa}_{\text{SCH}}$. The conditional variance component ratios were computed as $\hat{\theta}_{\text{NEIGH}} = \hat{\tau}_{\text{NEIGH}}/\hat{\sigma}^2 = .0760$ and $\theta_{\text{SCH}} = \hat{\tau}_{\text{SCH}}/\hat{\sigma}^2 = .0536$, which did not show a substantial difference. However, with 524 neighborhoods and 17 schools, one gets $\hat{\kappa}_{\text{NEIGH}} = 2310/524 \times .0760 = 0.335$ and $\hat{\kappa}_{\text{SCH}} = 2310/17 \times .0536 = 7.285$ so accounting for school-level clustering might be more important even for neighborhood-level variable. This is consistent with the previous discussion that, when the number of clusters in neighborhood is large and the cluster size is small, its impact on the fixed effect standard error tends to zero. The Cramér's *V* for the data was computed as .875, so part of the missing cluster information when one of the crossed random effects is omitted may be recovered by the inclusion of the other level.

To compute the corrected *SE*, first note that the variance proportion of `DEPRIVE` at the school level (i.e., $f_{X\text{SCH}}$) was .152, and $\varphi^2 s_1 = V_c^2(b-1)/(a-1) = 0.0234$, so $\varphi^2 s_1 - f_{X\text{SCH}} < 0$, and so the *SE* for `DEPRIVE` will be underestimated based on our discussion on Level-*A* predictors (i.e., at the modeled crossed random effect). Under HLM-N, $\tilde{\kappa}_{\text{NEIGH}} = (2310/524) \times (.087/.812) = 0.4723$. Using equation (15), one obtained a relative bias of $\widehat{\text{Var}}(\tilde{\gamma}_{01})$ as $-0.38$, or a correction factor on the estimated *SE* of $\sqrt{1/(1-0.38)} = 1.27$, so the corrected *SE* under HLM-N is 0.0491, which was comparable to the Monte Carlo *SE* of 0.0475, even though the cluster sizes were quite unbalanced. Similarly, using equation (13) and $\tilde{\kappa}_{\text{SCH}} = (2310/17) \times (.046/.865) = 0.234$, one obtained a corrected *SE* under HLM-S of 0.0367, which was also slightly underestimated but comparable to the Monte Carlo *SE* of 0.0387. Therefore, using the bias expressions presented in this paper, one can obtain a less biased estimate of the fixed effect *SE*s if an original study ignored either the neighborhood-level or the school-level clustering.

## Correction When True Parameters Are Unknown

The previous computation of the corrected *SE* when neighborhood was omitted (i.e., under HLM-S) assumed that $\kappa_{\text{NEIGH}}$, Cramér's $V$, and $f_{X\text{SCH}}$, were known. In reality, however, these quantities are most likely unknown, and educated guess are needed based on previous literature as well as researchers' substantive knowledge. For our example, Brännström (2008) reported that the intraclass correlation of mathematics achievement with respect to neighborhood was .084 based on a sample of 26,384 students in Sweden secondary schools, which can be converted to $\hat{\theta}_{\text{NEIGH}} = .084/(1 - .084) = .092$, and so $\hat{\kappa}_{\text{NEIGH}} = 2310/524 \times .092 = 0.40$. For Cramér's $V$, as previous studies generally did not report its value for the cross-classification of neighborhood and school, one needs to make an educated guess. As students in the same neighborhood tend to attend the same school, we expect $V_c$ to be large. To be conservative we assume that $V_c^2 = .50$, as a smaller $V_c$ gives a larger corrected *SE*. Finally, we approximated $f_{X\text{SCH}}$ as $\varphi^2 s_1 = 0.5$, which approximates its expected value for a stochastic neighborhood-level predictor across schools, and $f_{xB}^O = 1 + \tilde{\kappa}_{\text{SCH}}(1 - f_{X\text{SCH}})^2 = 20.02$. Using equation (13), one estimates the relative bias as $-0.1169$, and so the corrected $SE = .033/(1 - 0.1169) = 0.037$, which was close to the true *SE* of $\tilde{\gamma}_{01}$ under HLM-S. As a smaller $f_{X\text{SCH}}$ results in a bigger correction on the *SE*, because it implies that less information is captured for the school level, one can obtain a more conservative corrected *SE* by using $f_{X\text{SCH}} = \varphi^2 s_1/2 = 0.25$, which resulted in a relative bias estimate of $-0.1728$ and a corrected *SE* of 0.040, slightly larger than the true *SE* of $\tilde{\gamma}_{01}$ under HLM-S.

## Discussion and Conclusion

In the present paper, we discussed methods to quantify the degree of unbalanced cell sizes in cross-classified data, and provided analytical expressions of the relative bias on the sampling variance of fixed effects when one of the two crossed random effects is ignored. The discussion has both theoretical and practical implications. Theoretically, to our knowledge our paper is the first in the multilevel modeling literature to discuss the associations between two crossed random effects of clusterings and link statistical indices in contingency table literature (e.g., Cramér's $V$) to discuss clustering structures in multilevel data. With that, we were able to extend the analytical and simulation results from Luo and Kwok (2009) and Meyers and Beretvas (2006) to a more general framework for evaluating the impact of omitting a crossed random effect for unbalanced

Table 1
*Multilevel Results for Scotland Neighborhood Study*

| | HLM-N | HLM-S | CCREM |
|---|---|---|---|
| Intercept | $0.100^*$ | 0.100 | 0.095 |
| | (0.024) | (0.056) | (0.056) |
| DEPRIVE | $-0.521^*$ | $-0.473^*$ | $-0.466^*$ |
| | (0.038) | (0.033) | (0.0383) |
| Corrected *SE* | (0.0491) | (0.0367) | — |
| Monte Carlo *SE* | (0.0475) | (0.0387) | (0.0382) |
| AIC | 6283.9 | 6269.6 | 6251.9 |
| BIC | 6306.9 | 6292.6 | 6280.6 |
| $\hat{\tau}_{\text{NEI}}$ | 0.087 | | 0.061 |
| $\hat{\tau}_{\text{SCH}}$ | | 0.046 | 0.043 |
| $\hat{\sigma}^2$ | 0.812 | 0.865 | 0.807 |

*Note.* HLM-N = Two-level HLM accounting only for the neighborhood clustering effect. HLM-S = Two-level HLM accounting only for the school clustering effect. CCREM = Cross-classified random effect model accounting for both the neighborhood and the school clustering effects. DEPRIVE = neighborhood-level social deprivation. The outcome is student-level total attainment. Standard errors are shown in parentheses.
$^*p < .001$.

CCREMs, where we showed that the bias mainly depends on (a) the clustering effects of the two levels (i.e., $\kappa_A$ and $\kappa_B$), (b) the association between them (i.e., $V_c$), and (c) the proportions of sum of squares of the predictor at the two crossed random effects (i.e., $f_{xA}$ and $f_{xB}$) and at the within-cell level. As recent literature provided more attentions to clustered data with a non-hierarchical structure, such as a cross-classified structure, more methodological work is needed to understand the theoretical properties of the corresponding modeling approaches.

The findings from previous literature were also mostly consistent with our framework. For example, Meyers and Beretvas (2006) found that the magnitude of the downward *SE* bias on the fixed effect estimates at the omitted level was larger with larger intraclass correlation and larger cluster size of the omitted level, but was smaller when the residuals of the two crossed random effects (i.e., $\mathbf{u}$ and $\mathbf{v}$ in our notations) were correlated. It is clear that larger intraclass correlation and cluster size contributes to $\kappa_B$, which has a large impact of the relative bias based on our analytical expression. The effect of correlated $\mathbf{u}$ and $\mathbf{v}$ can be explained as contributing to the association between $\mathbf{Z}_1\mathbf{u}$ and $\mathbf{Z}_2\mathbf{v}$ in the CCREM in equation (1) even under a balanced design, so the consequence is similar to unbalanced cell sizes which also leads to an association between $\mathbf{Z}_1\mathbf{u}$ and $\mathbf{Z}_2\mathbf{v}$ even when $\mathbf{u}$ and $\mathbf{v}$ are independent.

The results in the present paper also have practical implications for researchers, and perhaps are most important for evaluation and meta-analysis of primary studies that did not account for some crossed random effects of clustering. Given that CCREMs have only become popular in the past two decades, some empirical studies using multilevel analysis, especially the ones dated earlier, might have omitted one or more crossed random effects. As demonstrated in the real data illustration, using the analytical results in the present study, researchers can obtain a corrected standard errors/variances of the fixed effect estimates of interest, and thus can correct for the biases in statistical inference in previous studies, an idea emphasized in Hedges (2007b) and Hedges (2009) for studies with two-level hierarchical data. Despite the increasing amount of cluster-randomized trials in education and other social sciences in recent years, there have been relatively few studies that actively modeled crossed random effects, meaning that results of many treatment effects in the literature may need to be adjusted. For example, Moerbeek and Safarkhani (2018) provided an example where soldiers are randomly assigned to different therapeutic approaches with different therapists, but the soldiers are also naturally nested within army units. Another common example in education is when classrooms of students are randomly assigned, but students are cross-classified by current classroom (e.g., 2nd grade) and also past-year classroom (e.g., 1st grade), yet primary studies may not have accounted for the shared variance of Grade 1 classroom/teacher effects. The corrections presented in this paper are thus crucial for evaluating whether conclusions in primary studies may change if previously omitted level of clustering had been appropriately accounted for. They are also needed for meta-analyses when some primary studies to be synthesized have complex data structures, but some crossed random effect had been omitted.

To successfully obtain a corrected variance or *SE* estimate when evaluating a primary study or doing a meta-analysis using the results in this paper, one needs to have some information on the clustering effects ($\kappa_A$ and $\kappa_B$) of the crossed random effects as well as the associations between the levels in observational data. Whereas one can easily obtain the clustering effects from the intraclass correlations, the values of which have been well researched for different educational outcomes on varying grade levels and demographic groups (Hedges & Hedberg, 2007; Kelcey & Phelps, 2013), very little is known about the association between crossed random effects, such as middle school and high school, school and neighborhood, and kindergarten classroom and Grade 1 classroom. Therefore, we recommend researchers doing primary research using CCREM report the Cramér's $V$ value of their crossed random effects, just like it is common practice to report the intraclass correlation in HLM. In addition, for both HLM and CCREM it is useful to report the proportion of variance of the predictors at different levels (e.g., $f_{xA}$ and $f_{xB}$ for CCREM), as they helped correct for the biases of ignoring a level of clustering in previous studies (see also Berkhof & Kampen, 2004). We also encourage future systematic reviews on the associations between crossed random effects commonly observed in CCREM. These not only help facilitate successful corrections in meta-analysis, but also potentially lead to benchmark values for design parameters when designing studies involving cross-classified data when the researchers do not have total control over the degree of imbalance of cell sizes.

The results also have practical implications for researchers designing primary studies that involve multiple crossed random effects of clustering. When designing such studies, whereas researchers usually have some control over the primary sampling unit, such as the number of schools and the number of students in each school, they generally do not have control on the other crossed random effect, such as the number of neighborhoods and the cell sizes for the cross-classification of schools and neighborhoods. Understanding the impact of each crossed random effect on the fixed effect variance estimates thus helps researchers determine which level(s) are the most important to have membership information collected. For example, from our real data illustration it was shown that, even though deprivation is a neighborhood level variable, one needs to collect information on school-level clustering and use an appropriate CCREM, because of the strong design effect (i.e., high intraclass correlation and/or large cluster size) of the school level and the high variability of deprivation across different schools (i.e., high $f_{X\text{SCH}}$).

The present study provided analytical expressions on factors affecting the relative bias induced by omitting a crossed random effect in CCREMs, and illustrated their usage in a real data example. We believe that this will help both substantive researchers and quantitative methodologists better understand these models. The limitations of the present study include the focus on models with only two crossed random effects and the assumption that there is no interaction between crossed random effects and no random slopes in the data and the model.[2] Also, our derivation was based on the assumption that the cluster sizes at each of the two crossed random effects are constant, and even though a small follow-up simulation showed the corrections still work to correct most of the bias, they may be less effective with highly unbalanced cluster sizes, or other combinations of Cramér's *V* and unbalanced cluster sizes. Future research effort is needed to address these issues.

# References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Berkhof, J., & Kampen, J. K. (2004). Asymptotic effect of misspecification in the random part of the multilevel model. *Journal of Educational and Behavioral Statistics*, *29*(2), 201–218. https://doi.org/10.3102/10769986029002201

Brännström, L. (2008, may). Making their mark: The effects of neighbourhood and upper secondary school on educational achievement. *European Sociological Review*, *24*(4), 463–478. Retrieved from `https://academic.oup.com/esr/article-lookup/doi/10.1093/esr/jcn013` https://doi.org/10.1093/esr/jcn013

Garner, C. L., & Raudenbush, S. W. (1991). Neighborhood effects on educational attainment: A multilevel analysis. *Sociology of Education*, *64*, 251–262. https://doi.org/10.2307/2112706

Goldstein, H. (1994). Multilevel cross-classified models. *Sociological Methods & Research*, *22*(3), 364–375. https://doi.org/10.1177/0049124194022003005

Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications* (2nd ed.). New York, NY: Routledge.

Hedges, L. V. (2007a). Correcting a significance test for clustering. *Journal of Educational and Behavioral Statistics*, *32*, 151–179. https://doi.org/10.3102/1076998606298040

Hedges, L. V. (2007b). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*, 341–370. https://doi.org/10.3102/1076998606298043

Hedges, L. V. (2009). Adjusting a significance test for clustering in designs with two levels of nesting. *Journal of Educational and Behavioral Statistics*, *34*, 464–490. https://doi.org/10.3102/1076998609337251

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*, 60–87. https://doi.org/10.3102/0162373707299706

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). New York, NY: Routledge.

Ingles, S. J., Abraham, S. Y., Karr, R., Spencer, B. D., & Frankel, M. R. (1990). *National Educational Longitudinal Study of 1988: Base year student component data file user's manual.* Washington, DC: U.S. Department of Education.

Kelcey, B., & Phelps, G. (2013). Considerations for designing group randomized trials of professional development with teacher knowledge outcomes. *Educational Evaluation and Policy Analysis*, *35*(3), 370–390. https://doi.org/10.3102/0162373713482766

Littell, R. C. (2006). *SAS for mixed models*. Cary, NC: SAS Institute.

Luo, W., & Kwok, O.-m. (2009). The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivariate Behavioral Research*, *44*, 182–212. https://doi.org/10.1080/00273170902794214

Luo, W., & Kwok, O.-m. (2012). The consequences of ignoring individuals' mobility in multilevel growth models: A Monte Carlo study. *Journal of Educational and Behavioral Statistics*, *37*(1), 31–56. https://doi.org/10.3102/1076998610394366

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*, 86–92. https://doi.org/10.1027/1614-1881.1.3.86

Meyers, J. L., & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research*, *41*(4), 473–497. https://doi.org/10.1207/s15327906mbr4104_3

Moerbeek, M. (2004). The consequence of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research*, *39*(1), 129–149. https://doi.org/10.1207/s15327906mbr3901_5

Moerbeek, M., & Safarkhani, M. (2018, apr). The design of cluster randomized trials with random cross-classifications. *Journal of Educational and Behavioral Statistics*, *43*(2), 159–181. Retrieved from `http://journals.sagepub.com/doi/10.3102/1076998617730303` https://doi.org/10.3102/1076998617730303

Norusis, M. J. M. J. (2012). *IBM SPSS statistics 19 advanced statistical procedures companion*. Upper Saddle River, NJ: Prentice Hall.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine: Series 5*, *50*(302), 157–175. https://doi.org/10.1080/14786440009463897

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2018). nlme: Linear and nonlinear mixed effects models [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=nlme` (R package version 3.1-137)

Rasbash, J., Steele, F., Browne, W. J., & Goldstein, H. (2017). A user's guide to MLwiN, v3.01 [Computer software manual]. Retrieved from `http://www.bristol.ac.uk/cmm/media/software/mlwin/downloads/manuals/3-01/manual-web.pdf`

Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data With applications in cross-sectional and longitudinal research. *Journal of Educational and Behavioral Statistics*, *18*(4), 321–349. https://doi.org/10.3102/10769986018004321

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

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & Du Toit, M. (2011). HLM 7: Hierarchical linear and nonlinear modeling [Computer software manual]. Lincolnwood, IL: Scientific Software International.

Searle, S. R., Casella, G., & McCulloch, C. E. (2006). *Variance components* (2nd ed.). Hoboken, NJ: Wiley.

Shi, Y., Leite, W., & Algina, J. (2010). The impact of omitting the interaction between crossed factors in cross-classified random effects modelling. *British Journal of Mathematical and Statistical Psychology*, *63*, 1–15. https://doi.org/10.1348/000711008X398968

Snijders, T. A. B. (2005). Power and sample size in multilevel modeling. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol. 3, pp.

1570–1573). Chicester, UK: Wiley. Retrieved from `http://www.stats.ox.ac.uk/~snijders/PowerSampleSizeMultilevel.pdf`

Snijders, T. A. B., & Bosker, R. J. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics*, *18*, 237–259. https://doi.org/10.3102/10769986018003237

Van den Noortgate, W., Opdenakker, M.-C., & Onghena, P. (2005). The effects of ignoring a level in multilevel analysis. *School Effectiveness and School Improvement*, *16*(3), 281–303. https://doi.org/10.1080/09243450500114850

Wallace, M. L. (2015). *Modeling cross- classified data with and without the crossed factors' random effects' Interaction* (Doctoral dissertation, University of Texas at Austin). Retrieved from `https://repositories.lib.utexas.edu/handle/2152/31010`

# Footnotes

[1] It is rare in real research, specially for observational studies, to have a level-1 predictor with identical cell means so that $f_{xA}$ and $f_{xB}$ are zero exactly, which means that all the cell means are exactly the same. An example of such predictors is a level-1 binary treatment or intervention indicator with equal number of participants in the treatment and in the control condition for every clusters in $A$ and every clusters in $B$. For such a purely level-1 predictor, the relative bias expression simplifies to

$$\text{RBias}[\hat{\text{Var}}(\tilde{\gamma}^{(1)})] = (1 - \varphi'^2)s_2\kappa_B.$$

From the above expression, it is obvious that the relative bias is non-negative and is zero only when $\varphi'^2 = 1$ (i.e., $A$ nested within $B$) or when $\tau_B^2 = \kappa_B = 0$, which is consistent to the conclusion in Luo and Kwok (2009) that ignoring an upper-level crossed random effect resulted in overestimated sampling variance of the fixed effects at the next lower level for a balanced design. The limiting value for the relative bias is $[(b-1)/b](1-\varphi'^2)\theta_B$ as $c_B \to \infty$, so the upward bias is larger for balanced designs (i.e., $\varphi'^2$ is close to 0) and when the omitted level has a relatively large variance component.

[2] We re-analyzed the data in the Scotland Neighborhood Study, and despite a statistically significant interaction effect between the crossed random effects on the intercept-only model, $\hat{\tau}_{A\times B}^2 = .068$, $p = .021$, the Monte Carlo standard errors of the misspecified fixed effect estimators changed by only 1% when the data generating model includes the interaction: $\text{Var}[\tilde{\gamma}] = 0.0481$ [vs. 0.0475 in Table 1] under HLM-N and 0.0392 [vs. 0.0387 in Table 1] under HLM-S.

# Appendix A

## Variance Component Estimators Under the Misspecified Model

Under the cross-classified model defined in (1), using the results from Searle, Casella, and McCulloch (2006, p. 183), the expected sum of squares (E[$SS$]) of the outcome $Y$ can be partitioned as

$$\begin{aligned} \mathrm{E}(SS_A) &= (N - k_1)\tau_A^2 + (k_3 - k_2)\tau_B^2 + (a-1)\sigma^2 \\ \mathrm{E}(SS_B) &= (k_4 - k_1)\tau_A^2 + (N - k_2)\tau_B^2 + (b-1)\sigma^2 \\ \mathrm{E}(SS_e) &= (k_1 - k_4)\tau_A^2 + (k_2 - k_3)\tau_B^2 + (N - a - b + 1)\sigma^2, \end{aligned}$$

with degrees of freedom ($df$) = $a - 1$ for $SS_A$, $b - 1$ for $SS_B$, and $N - a - b + 1$ for $SS_e$, and

$$k_1 = \sum_i n_{i.}^2/N, \tag{A1}$$

$$k_2 = \sum_{j=1}^{b} n_{.j}^2/N \tag{A2}$$

$$k_3 = \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}^2/n_{i.}, \tag{A3}$$

$$k_4 = \sum_{j=1}^{b}\sum_{i=1}^{a} n_{ij}^2/n_{.j} \tag{A4}$$

With equal cluster sizes for $A$ and $B$ such that $n_{1.} = n_{2.} = \ldots = n_{a.} = c_A$ and $n_{.1} = n_{.2} = \ldots = n_{.b} = c_B$ , we have $k_2 = c_B$ and $k_3 = \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}^2/c_A$. However, when level $B$ is omitted, we have $\mathrm{E}(SS_e^*) = \mathrm{E}(SS_B) + \mathrm{E}(SS_e)$, which equals

$$\mathrm{E}(SS_e^*) = (N - k_3)\tau_B^2 + (N - a)\sigma^2$$

with $df = N - a$. One way to measure the overlap between $A$ and $B$ is the ratio $(k_3 - k_2)/(N - k_2)$, which summarizes the relative influence of $\tau_B^2$ on $SS_A$ and $SS_B$ due to the unbalanced cell sizes. With equal cluster sizes for $B$, this ratio is simplified to $(k_3 - c_B)/[c_B(b - 1)]$ and can be shown equal to $\varphi'^2$ in equation (10).

When cell sizes are all equal to $\bar{n}$, $k_3 = ab\bar{n}^2/c_A = a\bar{n} = c_B$, so $k_3 - k_2 = c_B - c_B = 0$. As a result, $\varphi^2 = 0$, $\mathrm{E}(SS_A)$ is not a function of $\tau_B^2$, and all sum of squares in $B$ will go to the within-cluster level in the misspecified model such that $\mathrm{E}(SS_e^*) = c_B(b-1)\tau_B^2 + (N-a)\sigma^2$, and $\mathrm{E}(SS_A)$ is not a function of $\tau_B^2$; otherwise, $k_3 - k_2 > 0$, $\varphi > 0$, and $\mathrm{E}(SS_A)$ included a component due to $B$. It can be shown that the maximum of $k_3 - k_2$ is $N - c_B$ when $k_3 = N$, a situation where $\varphi'^2 = 1$ so that all sum of squares in the omitted level $B$ will go to $A$ (and $N - k_3 = 0$ so that $\mathrm{E}[SS_e^*]$ is not a function of $\tau_B^2$).

When level $B$ is ignored one obtains a balanced two-level model where the ANOVA estimates and the REML estimates of the variance components coincide (Searle et al., 2006). Using the results in Searle et al. (2006), the expected mean squares with equal cluster sizes can be expressed as

$$\begin{aligned}\mathrm{E}(MS_A) &= \frac{SS_A}{a-1} = c_A\tau_A^2 + \frac{k_3-k_2}{a-1}\tau_B^2 + \sigma^2\\ \mathrm{E}(MS_e^*) &= \frac{SS_e^*}{N-a} = \frac{N-k_3}{N-a}\tau_B^2 + \sigma^2.\end{aligned}$$

The estimated variance components under the misspecified model are

$$\begin{aligned}\tilde{\tau}_A^2 &= \frac{\mathrm{E}(MS_A) - \mathrm{E}(MS_e^*)}{c_A}\\ &= \tau_A^2 + \frac{[(k_3-k_2)/(a-1) - (N-k_3)/(N-a)]\tau_B^2}{c_A}\\ &= \tau_A^2 + \frac{c_B}{c_A}[\varphi^2 s_1 - (1-\varphi'^2)s_2]\tau_B^2\\ &= \frac{\sigma^2}{c_A}[\kappa_A + [\varphi^2 s_1 - (1-\varphi'^2)s_2]\kappa_B],\end{aligned}$$

and

$$\begin{aligned}\tilde{\sigma}^2 &= \mathrm{E}(MS_e^*) = \sigma^2 + \tau_B^2 \frac{N - k_3}{N - a} \\ &= \sigma^2 + c_B(1 - \varphi'^2)s_2\tau_B^2 \\ &= \sigma^2[1 + (1 - \varphi'^2)s_2\kappa_B],\end{aligned}$$

where $s_1 = 1/(a - 1)$ and $s_2 = (b - 1)/(N - a)$.

# Appendix B

## True Sampling Variance of the Fixed Effect Estimators Under a Misspecified Model

When cluster sizes in level $A$ are equal such that $\mathbf{Z}_1 = \mathbf{I}_a \otimes \mathbf{1}_{c_A}$, it can be shown that $\tilde{\mathbf{V}}^{-1} = (\sigma^{-2}) \left( \mathbf{I}_N - \frac{\tilde{\theta}_A}{1+\tilde{\kappa}_A} \mathbf{Z}_1 \mathbf{Z}_1' \right)$. From equation (5), if we let $\mathbf{Q}_1 = \mathbf{x}'\tilde{\mathbf{V}}^{-1}\tilde{\mathbf{V}}^{-1}\mathbf{x}$, $\mathbf{Q}_2 = \mathbf{x}'\tilde{\mathbf{V}}^{-1}\mathbf{Z}_1\mathbf{Z}_1'\tilde{\mathbf{V}}^{-1}\mathbf{x}$, and $\mathbf{Q}_3 = \mathbf{x}'\tilde{\mathbf{V}}^{-1}\mathbf{Z}_2\mathbf{Z}_2'\tilde{\mathbf{V}}^{-1}\mathbf{x}$, then the relative bias can be expressed as

$$\mathrm{RBias}[\hat{\mathrm{Var}}(\tilde{\gamma})] = \frac{(\tilde{\sigma}^2 - \sigma^2)\mathbf{Q}_1 + (\tilde{\tau}_A^2 - \tau_A^2)\mathbf{Q}_2 - \tau_B^2 \mathbf{Q}_3}{\sigma^2 \mathbf{Q}_1 + \tau_A^2 \mathbf{Q}_2 + \tau_B^2 \mathbf{Q}_3}, \tag{A5}$$

so by finding $\mathbf{Q}_1$, $\mathbf{Q}_2$, $\mathbf{Q}_3$ and substituting them into the expression will lead to equation (13). With some algebra,

$$\begin{aligned}
\mathbf{Q}_1 &= \frac{SS_x}{\tilde{\sigma}^4} \left[ (1 - f_{xA}) + \frac{f_{xA}}{(1+\tilde{\kappa}_A)^2} \right] \\
\mathbf{Q}_2 &= \frac{SS_x}{\tilde{\sigma}^4} \frac{c_A f_{xA}}{(1+\tilde{\kappa}_A)^2} \\
\mathbf{Q}_3 &= \frac{1}{\tilde{\sigma}^4} \left[ \mathbf{x}'\mathbf{Z}_2\mathbf{Z}_2'\mathbf{x} - \frac{2\tilde{\theta}_A}{1+\tilde{\kappa}_A} \mathbf{x}'\mathbf{Z}_1\mathbf{Z}_1'\mathbf{Z}_2\mathbf{Z}_2'\mathbf{x} + \frac{\tilde{\theta}_A^2}{(1+\tilde{\kappa}_A)^2} \mathbf{x}'\mathbf{Z}_1\mathbf{Z}_1'\mathbf{Z}_2\mathbf{Z}_2'\mathbf{Z}_1\mathbf{Z}_1'\mathbf{x} \right] \\
&= \frac{SS_x}{\tilde{\sigma}^4} \left[ c_B f_{xB} - \frac{2\tilde{\theta}_A}{1+\tilde{\kappa}_A} c_A c_B f_{xA} f_{xB} + \frac{\tilde{\theta}_A^2}{(1+\tilde{\kappa}_A)^2} \frac{c_A^2 \mathbf{x}_A'\mathbf{Z}_2\mathbf{Z}_2'\mathbf{x}_A}{SS_x} \right] \\
&= \frac{SS_x}{\tilde{\sigma}^4} \left[ c_B f_{xB} - \frac{2c_B\tilde{\kappa}_A}{1+\tilde{\kappa}_A} f_{xA} f_{xB} + \frac{\tilde{\kappa}_A^2}{(1+\tilde{\kappa}_A)^2} \frac{\mathbf{x}_A'\mathbf{Z}_2\mathbf{Z}_2'\mathbf{x}_A}{SS_x} \right]
\end{aligned}$$

Also, using the expressions of $\tilde{\tau}_A^2$ and $\tilde{\sigma}^2$ in Appendix A, we have

$$(1+\tilde{\kappa}_A)\frac{\tilde{\sigma}^2}{\sigma^2} = \left( 1 + \frac{c_A \tilde{\tau}_A^2}{\tilde{\sigma}^2} \right) \frac{\tilde{\sigma}^2}{\sigma^2} = \left[ 1 + \frac{\mathrm{E}(MS_A) - \mathrm{E}(MS_e^*)}{\mathrm{E}(MS_e^*)} \right] \times \frac{\mathrm{E}(MS_e^*)}{\sigma^2} = 1 + \kappa_A + \varphi^2 s_1 \kappa_B.$$

Substituting the above expressions, together with those for $\tilde{\tau}_A^2$ and $\tilde{\sigma}^2$, into (A5), and further simplifying, one obtains (13).