MULTIVARIATE SMALL-AREA ESTIMATION WITH SHRINKAGE COVARIANCE MATRICES: REVIEW AND EMPIRICAL STUDY IN COMPARISON WITH LINEAR MIXED-EFFECTS MODELING

DARIUS LI

Bureau of Economic and Business Research
David Eccles School of Business
University of Utah

Abstract. Multivariate composite estimation involves leveraging sampling strength across both subpopulations and small areas. We review Longford's multivariate composite estimation method and conduct a simulation study to assess the improvement over the baseline univariate analog. We also examine the empirical conditions under which multivariate composite estimation yields improvements over the baseline univariate method. In estimating the covariance matrix term in the expected mean squared error of the multivariate composite form, we adopt the Ledoit-Wolf approach of shrinkage covariance matrix estimation rather than employ moment-matching estimators. The use of moment-matching estimators works fairly well in the univariate case, but often cannot yield positive semi-definite matrices for the multivariate form. In addition to modifying the multivariate composite estimation method, we compare the empirical results to a simple linear mixed-effects model with the subpopulation as the fixed effect and the district as random effect.

1. Review of Multivariate Composite Estimation

Longford has done extensive research on design-based composite estimators. His baseline univariate method involves leveraging sampling strength from large geographies to develop composite estimators that produce more reliable small-area estimates [5]. We have reviewed this method and conducted a preliminary application of the method using the American Community Survey PUMS data [2]. This baseline method, however, does not systematically leverage sampling strength across subpopulations. Thus, in developing small-area estimates for a particular subpopulation, we are unable to leverage sample data from other demographic categories using the univariate method. For small demographic groups, this

Date: September 8, 2014.

The author would like to thank Michael T. Hogue, Senior Research Statistician at the Bureau of Economic and Business Research, for his review of the statistical content in this working paper as well as John C. Downen, Senior Research Analyst at BEBR, for his editing of the manuscript. Any remaining errors are the author's own.

baseline univariate estimation method may not fully capture the potential sampling strength available in the entire survey data. Thus, Longford introduces an important multivariate extension that leverages sampling strength not only across geographical levels but also across subpopulations [4].

1.1. **Composite Form.** Longford's multivariate composite shrinkage estimator is of the following form:

$$\tilde{\theta}_d = \hat{\boldsymbol{\theta}}_d^T(\mathbf{w} - \mathbf{b}_d) + \hat{\boldsymbol{\theta}}^T \mathbf{b}_d \tag{1.1}$$

where $\hat{\boldsymbol{\theta}}_d = (\hat{\theta}_d^{(1)}, \ldots, \hat{\theta}_d^{(k)})^T$ is the direct estimator of parameters in district d with k subpopulations, $\hat{\boldsymbol{\theta}}$ is the vector of national estimators for the corresponding parameters, and $\mathbf{w}$ consists of the weights assigned to the parameters. The vector $\mathbf{w}$ could be estimated using similar methods or could be derived from other data sources such as censuses or reliable auxiliary data. Note that the subpopulations could be based on demographic categories such as gender, age, race, or ethnicity. The geographical hierarchy is similarly flexible. Districts can be any geographic division of the "nation," which is itself not limited to any particular geographic level.

Our goal is to find the optimal weight vector $\mathbf{b}_d^\star$ that minimizes the expected mean squared error of the multivariate composite $\tilde{\theta}_d$ with respect to the vector $\boldsymbol{\theta}_d$ of the actual parameters. This loss function is of the following form:

$$\mathrm{eMSE}\left\{\tilde{\theta}_d(\mathbf{b}_d); \boldsymbol{\theta}_d^T \mathbf{w}\right\} \tag{1.2}$$

$$= \mathrm{Var}\left(\tilde{\theta}_d\right) + \mathrm{E}_{\mathcal{D}}\left[(\tilde{\theta}_d - \boldsymbol{\theta}_d^T \mathbf{w})^2\right] \tag{1.3}$$

$$= \mathrm{Var}\left\{\hat{\boldsymbol{\theta}}_d^T(\mathbf{w} - \mathbf{b}_d) + \hat{\boldsymbol{\theta}}^T \mathbf{b}_d\right\} + \mathrm{E}_{\mathcal{D}}\left\{\left[\boldsymbol{\theta}_d^T(\mathbf{w} - \mathbf{b}_d) + \boldsymbol{\theta}^T \mathbf{b}_d - \boldsymbol{\theta}_d^T \mathbf{w}\right]^2\right\} \tag{1.4}$$

$$\begin{aligned} &= (\mathbf{w} - \mathbf{b}_d)^T \mathrm{Var}(\hat{\boldsymbol{\theta}}_d)(\mathbf{w} - \mathbf{b}_d) + \mathbf{b}_d^T \mathrm{Var}(\hat{\boldsymbol{\theta}})\mathbf{b}_d \\ &\qquad + 2\mathbf{b}_d^T \mathrm{Cov}(\hat{\boldsymbol{\theta}}_d, \hat{\boldsymbol{\theta}})(\mathbf{w} - \mathbf{b}) + \mathbf{b}_d^T \mathrm{E}_{\mathcal{D}}\left[(\boldsymbol{\theta}_d - \boldsymbol{\theta})(\boldsymbol{\theta}_d - \boldsymbol{\theta})^T\right]\mathbf{b}_d \end{aligned} \tag{1.5}$$

$$= (\mathbf{w} - \mathbf{b}_d)^T \mathrm{Var}(\hat{\boldsymbol{\theta}}_d)(\mathbf{w} - \mathbf{b}_d) + \mathbf{b}_d^T \mathrm{Var}(\hat{\boldsymbol{\theta}})\mathbf{b}_d + 2\mathbf{b}_d^T \mathbf{N}_d \mathbf{V}_d(\mathbf{w} - \mathbf{b}) + \mathbf{b}_d^T \boldsymbol{\Sigma} \mathbf{b}_d \tag{1.6}$$

$$= \mathbf{w}^T \mathbf{V}_d \mathbf{w} - 2\mathbf{b}_d^T \mathbf{V}_d \mathbf{w} + \mathbf{b}_d^T \mathbf{V}_d \mathbf{b}_d + \mathbf{b}_d^T\left[\boldsymbol{\Sigma} + \mathbf{V}\right]\mathbf{b}_d + 2\mathbf{b}_d^T \mathbf{N}_d \mathbf{V}_d \mathbf{w} - 2\mathbf{b}_d^T \mathbf{N}_d \mathbf{V}_d \mathbf{b}_d \tag{1.7}$$

where $\mathbf{N}_d = \frac{1}{n_+}\mathrm{diag}_i\left(n_d^{(i)}\right)$, $\mathbf{V}_d = \mathrm{diag}_i\left(v_d^{(i)}\right)$, $\mathbf{V} = \mathrm{Var}(\hat{\boldsymbol{\theta}}) = \mathrm{diag}_i\left(v^{(i)}\right)$, and $\boldsymbol{\Sigma} = \mathrm{E}_{\mathcal{D}}\left[(\boldsymbol{\theta}_d - \boldsymbol{\theta})(\boldsymbol{\theta}_d - \boldsymbol{\theta})^T\right]$ is the multivariate between-area covariance matrix [4]. Note that we can normalize $n_+ = \sum_{i=1}^k n_d^{(i)}$ to 1 so that $n_d^{(i)}$ simply represents the sampling weight of subpopulation i within district d.

1.2. **Minimizing Expected MSE.** We obtain the local minimum of (1.2), using matrix differentiation as follows:

$$\frac{\partial\, \text{eMSE}\left\{\tilde{\theta}_d(\mathbf{b}_d); \boldsymbol{\theta}_d^T \mathbf{w}\right\}}{\partial \mathbf{b}_d} =$$

$$-2\mathbf{V}_d\mathbf{w} + (\mathbf{V} + \mathbf{V}^T)\mathbf{b}_d + \left\{\boldsymbol{\Sigma} + \mathbf{V} + (\boldsymbol{\Sigma} + \mathbf{V})^T\right\}\mathbf{b}_d$$

$$+ 2\mathbf{N}_d\mathbf{V}_d\mathbf{w} - 2\left[\mathbf{N}_d\mathbf{V}_d + (\mathbf{N}_d\mathbf{V}_d)^T\right]\mathbf{b}_d = 0$$

$$-2\mathbf{V}_d\mathbf{w} + 2\mathbf{V}_d\mathbf{b}_d + 2\left[\boldsymbol{\Sigma} + \mathbf{V}\right]\mathbf{b}_d + 2\mathbf{N}_d\mathbf{V}_d\mathbf{w} - 2\left[\mathbf{N}_d\mathbf{V}_d + \mathbf{V}_d\mathbf{N}_d\right]\mathbf{b}_d = 0 \tag{1.8}$$

By solving (1.8) for $\mathbf{b}_d$, we obtain

$$\mathbf{b}_d^{\star} = \mathbf{D}_d^{-1}(\mathbf{I} - \mathbf{N}_d)\mathbf{V}_d\mathbf{w} \tag{1.9}$$

where $\mathbf{D}_d = \mathbf{V}_d + \mathbf{V} + \boldsymbol{\Sigma} - \mathbf{N}_d\mathbf{V}_d - \mathbf{V}_d\mathbf{N}_d = \mathbf{V}_d + \mathbf{V} + \boldsymbol{\Sigma} - 2\mathbf{N}_d\mathbf{V}_d$.

We then substitute (1.9) into (1.2) to obtain the minimum expected mean squared error:

$$\text{eMSE}\left\{\tilde{\theta}_d(\mathbf{b}_d^{\star}); \boldsymbol{\theta}_d^T\mathbf{w}\right\} \tag{1.10}$$

$$= \mathbf{w}^T\mathbf{V}_d\mathbf{w} - 2[\mathbf{D}_d^{-1}(\mathbf{I} - \mathbf{N}_d)\mathbf{V}_d\mathbf{w}]^T(\mathbf{I} - \mathbf{N}_d)\mathbf{V}_d\mathbf{w}$$
$$[\mathbf{D}_d^{-1}(\mathbf{I} - \mathbf{N}_d)\mathbf{V}_d\mathbf{w}]^T(\mathbf{V}_d + \boldsymbol{\Sigma} + \mathbf{V} - 2\mathbf{V}_d\mathbf{V}_d)[\mathbf{D}_d^{-1}(\mathbf{I} - \mathbf{N}_d)\mathbf{V}_d\mathbf{w}] \tag{1.11}$$

$$= \mathbf{w}^T\mathbf{V}_d\mathbf{w} - [\mathbf{w}^T\mathbf{V}_d\mathbf{D}_d^{-1}]\mathbf{D}_d[\mathbf{D}_d^{-1}(\mathbf{I} - \mathbf{N}_d)\mathbf{V}_d\mathbf{w}] \tag{1.12}$$

$$= \mathbf{w}^T\mathbf{V}_d\mathbf{w} - \mathbf{b}_d^{\star T}\mathbf{D}_d\mathbf{b}_d^{\star} \tag{1.13}$$

In order to calculate the composite estimator $\tilde{\boldsymbol{\theta}}_d$, we substitute (1.9) into the composite form in (1.1).

$$\tilde{\theta}_d = \hat{\boldsymbol{\theta}}_d^T(\mathbf{w} - \mathbf{b}_d^{\star}) + \hat{\boldsymbol{\theta}}^T\mathbf{b}_d^{\star} \tag{1.14}$$

$$= \hat{\boldsymbol{\theta}}_d^T[\mathbf{I} - \mathbf{D}_d^{-1}(\mathbf{I} - \mathbf{N}_d)\mathbf{V}_d]\mathbf{w} + \hat{\boldsymbol{\theta}}^T\mathbf{D}_d^{-1}(\mathbf{I} - \mathbf{N}_d)\mathbf{V}_d\mathbf{w} \tag{1.15}$$

$$= \left\{\hat{\boldsymbol{\theta}}_d^T[\mathbf{I} - \mathbf{D}_d^{-1}(\mathbf{I} - \mathbf{N}_d)\mathbf{V}_d] + \hat{\boldsymbol{\theta}}^T\mathbf{D}_d^{-1}(\mathbf{I} - \mathbf{N}_d)\mathbf{V}_d\right\}\mathbf{w} \tag{1.16}$$

$$= \left\{[\mathbf{I} - \mathbf{V}_d(\mathbf{I} - \mathbf{N}_d)\mathbf{D}_d^{-1}]\hat{\boldsymbol{\theta}}_d + \mathbf{V}_d(\mathbf{I} - \mathbf{N}_d)\mathbf{D}_d^{-1}\hat{\boldsymbol{\theta}}\right\}\mathbf{w} \tag{1.17}$$

Since $\tilde{\theta}_d = \tilde{\boldsymbol{\theta}}_d\mathbf{w}$, from (1.17), we can extract the vector form of the composite estimator evaluated at optimal shrinkage weights:

$$\tilde{\boldsymbol{\theta}}_d = [\mathbf{I} - \mathbf{V}_d(\mathbf{I} - \mathbf{N}_d)\mathbf{D}_d^{-1}]\hat{\boldsymbol{\theta}}_d + \mathbf{V}_d(\mathbf{I} - \mathbf{N}_d)\mathbf{D}_d^{-1}\hat{\boldsymbol{\theta}} \tag{1.18}$$

2. Covariance Matrix Estimation

Similar to the case in univariate composite estimation, the expected mean squared error still has a term that requires further estimation. In the multivariate case, this term is $\mathbf{\Sigma} = \mathrm{E}_{\mathcal{D}}\left[(\boldsymbol{\theta}_d - \boldsymbol{\theta})(\boldsymbol{\theta}_d - \boldsymbol{\theta})^T\right]$, which is a critical component for solving the optimal shrinkage vector $\mathbf{b}_d^\star$ in (1.9). Longford proposed moment-matching estimators for the diagonal entries of $\mathbf{\Sigma}$ and sample covariances for the off-diagonal entries [4]. In practice, however, this method can lead to estimated covariance matrices that are not positive semi-definite, potentially resulting in outlier composite estimates. Despite the issue of estimating $\mathbf{\Sigma}$, Longford's multivariate composite estimation does in fact yield reductions in MSE when we use the actual value of $\mathbf{\Sigma}$ from simulated population data. Thus, we seek to modify only the estimation of $\mathbf{\Sigma}$ in using Longford's multivariate composite estimation method.

In practice, we would not have population data to calculate the actual value $\mathbf{\Sigma}$. Thus, we need an estimator for $\mathbf{\Sigma}$ that performs well even under cases of small sample sizes. This issue is pervasive in many applied fields with high dimensional data. Ledoit and Wolf proposed a shrinkage method to estimate the covariance variance of stock returns, using a composite of a sample covariance matrix and a target covariance matrix derived from a statistical model on stock returns [3]. Schäfer and Strimmer expanded upon the work of Ledoit and Wolf by examining various target covariance matrices in order to analyze large-scale gene association networks [6]. In place of moment-matching estimators and sample covariances, we will estimate $\mathbf{\Sigma}$ by testing some of target covariance matrices suggested by Schäfer and Strimmer based on the Ledoit-Wolf approach of shrinkage covariance matrix estimation.

Schäfer and Strimmer explain that empirical sample covariance matrices are often not appropriate for estimating the actual covariance matrix, especially in cases of small sample sizes [6]. Similarly, Ledoit and Wolf note that while the sample covariance matrix has appealing properties of being the maximum likelihood estimate under normality, this only implies that the sample covariance matrix will perform well asymptotically as the number of observations approaches infinity [3]. Thus, with small samples, the sample covariance matrix may not perform well as an estimator.

Hence, we modify Longford's multivariate composite estimatoin method by adopting the Ledoit-Wolf approach of estimating covariance matrices. Under this framework, the shrinkage covariance matrix $\hat{\mathbf{\Sigma}}$ is a weighted average of the target covariance matrix $\mathbf{T}$ and the sample covariance matrix $\mathbf{S}$ in the following form

$$\hat{\mathbf{\Sigma}} = \lambda \mathbf{T} + (1-\lambda)\mathbf{S} \tag{2.1}$$

where $\lambda \in [0,1]$ is the shrinkage weight assigned to the target covariance matrix.

Ledoit and Wolf employ the squared Frobenius norm

$$L(\lambda) = \left\| \lambda \mathbf{T} + (1-\lambda)\mathbf{S} - \mathbf{\Sigma} \right\|_F^2 \tag{2.2}$$

as the quadratic loss function to solve for the optimal $\lambda^\star$ that minimizes $L(\lambda)$. The resulting $\hat{\mathbf{\Sigma}}(\lambda^\star)$ is the optimal shrinkage covariance matrix [3].

Of the six commonly used target covariance matrices described by Schäfer and Strimmer, we will focus on four. The first target covariance $\mathbf{T}_A$ consists of unit variances, where

$$(\mathbf{T}_A)_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{2.3}$$

The following shrinkage weight $\lambda_A^\star$ yields the optimal shrinkage covariance matrix using $\mathbf{T}_A$ as the target covariance matrix:

$$\hat{\lambda}_A^\star = \frac{\sum_{i\neq j} \widehat{\text{Var}}\left(s_{ij}\right) + \sum_i \widehat{\text{Var}}\left(s_{ii}\right)}{\sum_{i\neq j} s_{ij}^2 + \sum_i (s_{ii} - 1)^2} \tag{2.4}$$

where s_{ij} is the element at the intersection of the ith row and jth column of the sample covariance matrix $\mathbf{S}$ [6].

The second target covariance matrix $\mathbf{T}_C$ consists of common variances and covariances in the following form:

$$(\mathbf{T}_C)_{ij} = \begin{cases} \text{avg}(s_{ii}) & \text{if } i = j \\ \text{avg}(s_{ij}) & \text{if } i \neq j \end{cases} \tag{2.5}$$

The following shrinkage weight optimizes the shrinkage covariance matrix estimator with the composite form of $\mathbf{T}_C$ and the sample covariance matrix:

$$\hat{\lambda}_C^\star = \frac{\sum_{i\neq j} \widehat{\text{Var}}(s_{ij}) + \sum_i \widehat{\text{Var}}(s_{ii})}{\sum_{i\neq j} (s_{ij} - c)^2 + \sum_i (s_{ii} - v)^2} \tag{2.6}$$

where $c = \text{avg}(s_{ij})$ and $v = \text{avg}(s_{ii})$ [6].

The third target covariance matrix $\mathbf{T}_D$ contains unequal variances along the diagonal entries in the following form:

$$(\mathbf{T}_D)_{ij} = \begin{cases} s_{ii} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{2.7}$$

The corresponding shrinkage weight

$$\hat{\lambda}_D^\star = \frac{\sum_{i\neq j} \widehat{\text{Var}}(s_{ij})}{\sum_{i\neq j} s_{ij}^2} \tag{2.8}$$

optimizes the shrinkage covariance matrix using the target $\mathbf{T}_D$ [6].

For highly correlated data, Schäfer and Strimmer suggest using this Target E covariance matrix $\mathbf{T}_E$, whose entries are defined as [6]

$$(\mathbf{T}_E)_{ij} = \begin{cases} s_{ii} & \text{if } i = j \\ \sqrt{s_{ii}s_{jj}} & \text{if } i \neq j \end{cases} \tag{2.9}$$

The optimal shrinkage weight corresponding to the target matrix $\mathbf{T}_E$ is

$$\hat{\lambda}_E^\star = \frac{\sum_{i\neq j} \widehat{\text{Var}}(s_{ij}) - f_{ij}}{\sum_{i\neq j} (s_{ij} - \sqrt{s_{ii}s_{jj}})^2} \tag{2.10}$$

where

$$f_{ij} = \frac{1}{2}\left\{\sqrt{\frac{s_{jj}}{s_{ii}}}\widehat{\text{Cov}}(s_{ii}, s_{ij}) + \sqrt{\frac{s_{ii}}{s_{jj}}}\widehat{\text{Cov}}(s_{jj}, s_{ij})\right\} \tag{2.11}$$

The term $\widehat{\text{Var}}\left(s_{ij}\right)$ found in (2.4), (2.6), (2.8), and (2.10) is the empirical unbiased variance of the entry at the intersection of the ith row and jth column of $\mathbf{S}$, which is defined in the following form [6]:

$$\begin{aligned}\widehat{\text{Var}}\left(s_{ij}\right) &= \widehat{\text{Var}}\left(\frac{D}{D-1}\bar{w}_{ij}\right) = \frac{D^2}{(D-1)^2}\widehat{\text{Var}}\left(\bar{w}_{ij}\right) = \frac{D}{(D-1)^2}\widehat{\text{Var}}\left(w_{ij}\right) \\ &= \frac{D}{(D-1)^3}\sum_{d=1}^{D}(w_{dij} - \bar{w}_{ij})^2\end{aligned} \tag{2.12}$$

where $w_{dij} = (\hat{\theta}_d^{(i)} - \bar{\hat{\theta}}_d^{(i)})(\hat{\theta}_d^{(j)} - \bar{\hat{\theta}}_d^{(j)})$, $\bar{w}_{ij} = \frac{1}{D}\sum_{d=1}^{D} w_{dij}$, and D is the number of districts that contain the subpopulations.

Similarly, $\widehat{\text{Cov}}\left(s_{ii}, s_{ij}\right)$ found in (2.11) is defined as

$$\widehat{\text{Cov}}(s_{ii}, s_{ij}) = \frac{D}{(D-1)^2}\widehat{\text{Cov}}(w_{ii}, w_{ij}) \tag{2.13}$$

Since the variance and covariance terms in (2.4), (2.6), (2.8), and (2.10) are replaced with unbiased sample variances and covariances, the estimated shrinkage weights $\hat{\lambda}_A^\star$, $\hat{\lambda}_C^\star$, $\hat{\lambda}_D^\star$, and $\hat{\lambda}_E^\star$ could lie outside the bounds of $[0, 1]$. As a remedy, we use the modified shrinkage weight $\hat{\lambda}_\Psi^\dagger = \max(0, \min(1, \hat{\lambda}_\Psi^\star))$ when calculating the shrinkage covariance matrix with the composite form of the target covariance matrix $\mathbf{T}_\Psi$ and the sample covariance matrix S [6].

As a result of attaining the modified shrinkage weight $\hat{\lambda}_\Psi^\dagger$, using some target covariance matrix $\mathbf{T}_\Psi$, we can calculate the estimated weights

$$\hat{\mathbf{b}}_d(\hat{\lambda}_\Psi^\dagger) = \hat{\mathbf{D}}_d^{-1}(\mathbf{I} - \mathbf{N}_d)\mathbf{V}_d\mathbf{w} \tag{2.14}$$

assigned to the multivariate composite estimator in (1.1), where $\hat{\mathbf{D}}_d = \mathbf{V}_d + \mathbf{V} + \hat{\mathbf{\Sigma}}(\hat{\lambda}_\Psi^\dagger) - \mathbf{N}_d\mathbf{V}_d - \mathbf{V}_d\mathbf{N}_d = \mathbf{V}_d + \mathbf{V} + \hat{\mathbf{\Sigma}}(\hat{\lambda}_\Psi^\dagger) - 2\mathbf{N}_d\mathbf{V}_d$. Note that (2.14) is the estimated analog of the optimal weight vector $\mathbf{b}_d^\star$ in (1.9).

3. Linear Mixed-Effects Model

Since the multivariate composite estimation method aims to improve estimates of subpopulations across districts, linear mixed-effects models can serve as a point of comparison. For the purposes of the simulation study in Section 4, we assign subpopulations as the fixed effects and districts as the random effects component of the model.

For comparability, we use the same sample data employed in the simulations for the multivariate composite estimation method as the input for the linear mixed-effects model. Since sample data consists of subpopulations across districts, we are modeling clustered data rather than longitudinal data. Specifically, the clusters in our case are subpopulations. Our response vector $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{id_i})^T$ for cluster i comprises of values of y_{ij}, where $i = 1, 2, \ldots, k$ and the index $j = 1, 2, \ldots, d_i$ refers to the repeated measures. In this case, k is the number of subpopulations and d_i is the number of districts with subpopulation i.

The following is the standard form of the linear mixed-effects model for cluster i [7]:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma} + \mathbf{e}_i \tag{3.1}$$

where $\mathbf{X}$ is the $d_i \times p$ fixed-effects design submatrix, $\boldsymbol{\beta}$ is the $p \times 1$ fixed-effects vector, $\mathbf{Z}_i$ is the $d_i \times q$ random-effects design submatrix, $\boldsymbol{\gamma}$ is the $q \times 1$ random-effects vector, and $\mathbf{e}_i$ represents the vector of random errors for within-cluster measurements. Note that p and q are the number of fixed and random effects, respectively.

The standard distributional assumptions of the model are $\boldsymbol{\gamma} \sim \mathcal{N}(0, \mathbf{D})$ and $\mathbf{e}_i \sim \mathcal{N}(0, \mathbf{R}_i)$, where $\mathbf{D}$ is the covariance matrix of the random effects and $\mathbf{R}_i$ is the covariance matrix of the measurements within subpopulation i.

4. Simulation Study

In this study, we simulate populations with two subpopulations (Groups X and Y) in five districts (A, B, C, D, and E). Thus, we construct a simple linear mixed-effects model[1] in which the subpopulations are the fixed effects and the districts are the random effects. This basic formulation is essentially a random intercept model. Since the multivariate small-area estimation method does not leverage additional data, we do not add complexity to the linear mixed-effects model for the purposes of comparability.

The simulations contain the following three major cases:

- DS: between-district dissimilarity of θ_d and similarity between subpopulations X and Y within each district
- DD: between-district and between-subpopulation dissimilarity
- SD: between-district similarity and between-subpopulation dissimilarity

Each of these three cases is paired with the following two different cases related to the underlying population standard deviation σ_d for each district d:

- e: equally large σ_d for both subpopulations across all districts d
- u: unequal σ_d split across subpopulations for all districts d

More specifically, in the equal-variance framework, the underlying population standard deviation for both subpopulations across all districts is roughly 10. On the other hand, in the unequal-variance framework, we specify σ_d to be roughly 6 and 10 for subpopulations X and Y, respectively, across all districts.

[1] We used the `lme4` package in `R` to model the simulated sample data [1].

The nomenclature of the simulation cases is a concatenation of the three major case names with the two cases related to the underlying population standard deviation. For instance, the simulated population for Case DDe has between-district and between-subpopulation dissimilarity with equally large σ_d for both subpopulations across all districts d.

Figures 1–6 show the empirical distributions of various estimators for each of the six simulation cases resulting from 1,000 repeated samples of the simulated populations. The sampling design of 2.5% simple random sampling without replacement within each subpopulation and each district remains consistent across all six simulation cases. Each district has a population size of 1,200 split evenly between both subpopulations X and Y. Note that we simulate different σ_d split across subpopulation lines, which is akin to testing different allocations of the populations between the two subpopulations. Thus, for the sake of simplicity and consistency, we do not alter the population allocations in the simulation cases.

For each simulation case, we compare the multivariate method, using various target covariance matrices, with the linear mixed-effects model and the univariate composite shrinkage estimator. In this simulation study, the univariate method leverages only the data within a given subpopulation. Since we examine the empirical results of the multivariate method, using target covariance matrices to estimate $\mathbf{\Sigma}$, we evaluate the univariate shrinkage estimator $\hat{\theta}_d^C$ at the estimated suboptimal weight $\hat{b}_\dagger$ for the purposes of comparability. Please see Economic and Statistical Working Paper No. 1 for a review of Longford's univariate composite estimation method [2].

From Tables 1–3, the empirical MSEs are listed for all six simulation cases. The notation of the estimators are simplified in the tables: $\tilde{\theta}_d^A$, $\tilde{\theta}_d^C$, $\tilde{\theta}_d^D$, and $\tilde{\theta}_d^E$ are the multivariate estimators using the target estimators $\mathbf{T}_A$, $\mathbf{T}_C$, $\mathbf{T}_D$, and $\mathbf{T}_E$, whereas $\tilde{\theta}_d^L$ is the predictor using the simple linear mixed-effects model. For graphical clarity, we omitted the empirical distributions of $\tilde{\theta}_d^A$ and $\tilde{\theta}_d^C$ from Figures 1–6.

4.1. **Case DS.** Figures 1 and 2 show the empirical distributions of estimators for Case DS, while Table 1 lists the empirical MSEs. Under Case DS, the actual parameters vary drastically across the five districts, whereas the subpopulations have very similar characteristics within each district. As a result, the univariate composite shrinkage estimator does not perform well, especially for Districts A and E in Case DSe (Figure 1), which exhibit incurred upward and downward bias, respectively, for the univariate method. In such cases, the univariate estimator performs worse than the original estimator $\hat{\theta}_d$ (Table 1). Since the univariate method can only leverage sample data across districts within a given subpopulation, between-district dissimilarity creates a huge handicap. The performance of the univariate estimator fares much better for Districts B, C, and D, since these areas have intermediate parameters values and are thus less prone to incurred bias from districts with more extreme values.

On the other hand, the multivariate estimators exhibit consistently better performance than the original estimator even for Districts A and E. The simulation results suggest that the multivariate estimators are less prone to incurred bias from dissimilar districts, since this method has the added flexibility of leveraging between subpopulations. This benefit is even more pronounced in Case DSu, in which the multivariate estimators further reduce the MSE for Group Y estimates by leveraging Group X survey data, which has a much smaller population standard deviation across districts than in Case DSe. Since the univariate method cannot leverage across subpopulations, the univariate shrinkage estimator cannot make any sizeable improvements in Group Y errors beyond those in Case DSe.

Of all the multivariate estimators tested in this simulation study, $\tilde{\theta}_d^C$ and $\tilde{\theta}_d^E$ tend for perform the best. This could be attributed to the off-diagonal entries in the target matrices $\mathbf{T}_C$ and $\mathbf{T}_E$, which are most appropriate for highly correlated data. The other target matrices have 0 in the off-diagonal matrices and probably do not suit correlated data as well, especially if the sample covariances are not reliable.

In fact, the multivariate estimators $\tilde{\theta}_d^C$ and $\tilde{\theta}_d^E$ perform almost as well as the linear mixed-effects model. For a few districts, linear mixed-effects model incurs more bias than $\tilde{\theta}_d^C$ and $\tilde{\theta}_d^E$. Thus, some of the multivariate estimators might be more favorable than the linear mixed-effects model in its simplest form if there is low tolerance for incurred bias.

4.2. **Case DD.** The underlying parameters differ not only across districts but also across subpopulations in Case DD (Figures 3 and 4). This difference from Case DS does not alter the behavior of the univariate shrinkage estimator, since this method does not leverage data across subpopulations. Similar to the results in Case DS, the multivariate estimators $\tilde{\theta}_d^C$ and $\tilde{\theta}_d^E$ yield smaller errors than those from $\tilde{\theta}_d^A$ and $\tilde{\theta}_d^D$, mostly due to the added covariance entries in the target matrices $\mathbf{T}_C$ and $\mathbf{T}_E$.

While the linear mixed-effects model yields the smallest errors for nearly all the districts, some sizeable biases are incurred for Districts A and E, most likely due to misspecified fixed and random effects from the linear predictor. For the multivariate estimators, however, the added feature of systematic differences in θ_d across subpopulations does not lead to any sizeable incurred bias. Thus, unless the random and fixed effects in the simple linear mixed-effects model are correctly specified across all the districts and subpopulations, multivariate estimators could still have an advantage in applications despite yielding slightly larger errors.

4.3. **Case SD.** Under the framework of Case SD, the underlying parameters are fairly similar across districts, but there exists a systematic difference in θ_d across subpopulations. Naturally, under such a data design, the univariate shrinkage estimator performs better than the set of mulivariate estimators (Table 3), since the systematic parameter difference across subpopulations renders the additional leveraging of data between Groups X and Y less useful. Nonetheless, the multivariate estimators do not exhibit any sizeable bias given the across-population dissimilarity. As with the previous cases, the linear mixed-effects model yields the smallest errors. While the linear mixed-effects model can specify fixed effects across subpopulations, the univariate method can only rely on the across-district similarity within a given subpopulation.

FIGURE 1. Case DSe

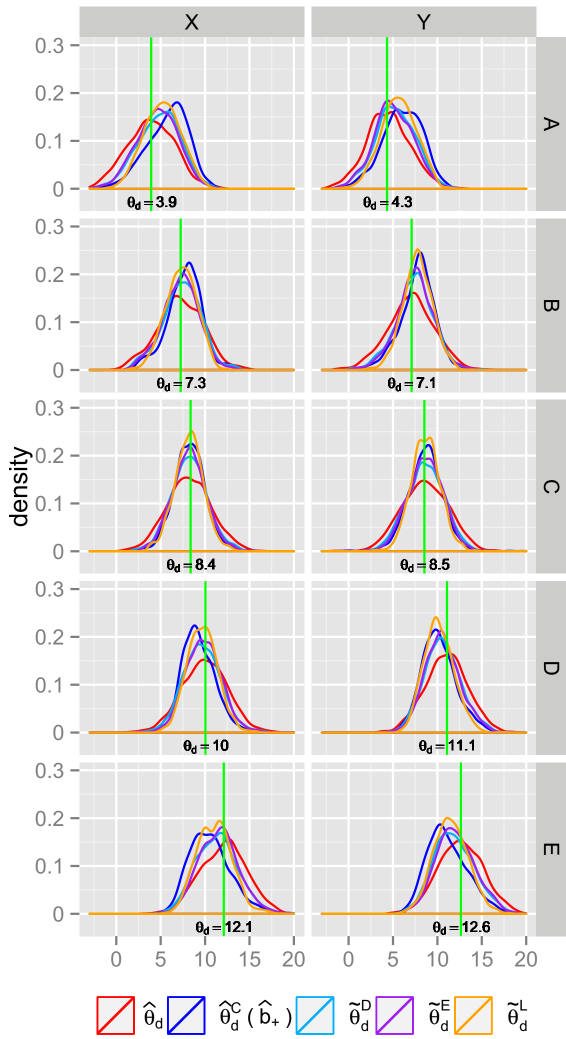

FIGURE 2. Case DSu

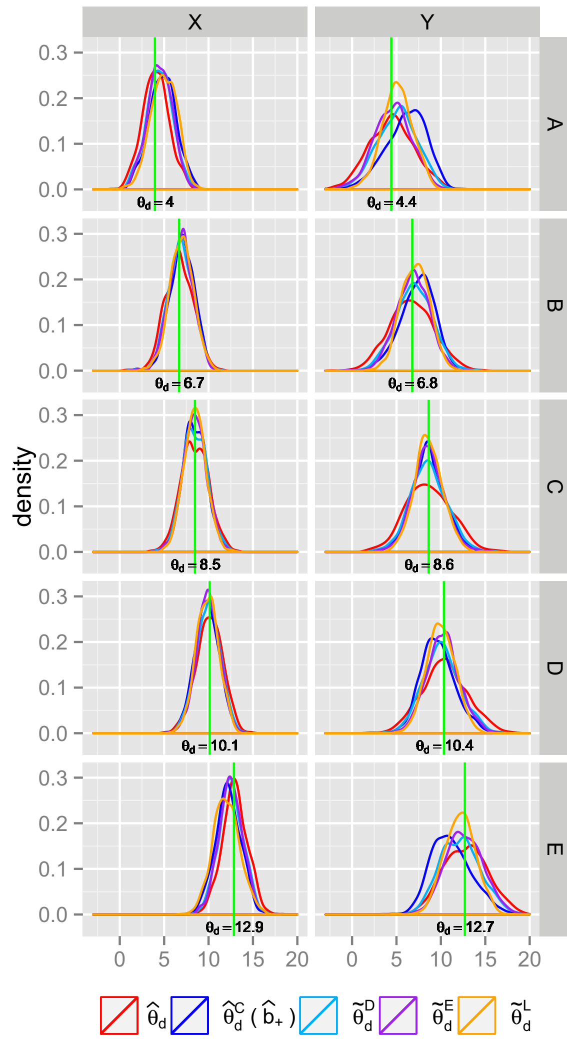

TABLE 1. Empirical MSE of Estimators for Case DS

			Case DSe							Case DSu						
		n_d	$\hat{\theta}_d$	$\hat{\theta}_d^C(\hat{b}_\dagger)$	$\tilde{\theta}_d^A$	$\tilde{\theta}_d^C$	$\tilde{\theta}_d^D$	$\tilde{\theta}_d^E$	$\tilde{\theta}_d^L$	$\hat{\theta}_d$	$\hat{\theta}_d^C(\hat{b}_\dagger)$	$\tilde{\theta}_d^A$	$\tilde{\theta}_d^C$	$\tilde{\theta}_d^D$	$\tilde{\theta}_d^E$	$\tilde{\theta}_d^L$
X	A	15	7.18	8.71	6.22	5.71	6.55	5.80	5.77	2.18	3.03	2.40	2.10	2.34	2.22	3.12
X	B	15	6.69	3.98	4.19	4.64	4.76	4.25	3.09	2.41	1.97	1.86	1.95	2.01	1.84	1.86
X	C	15	6.25	3.25	3.60	4.08	4.16	3.73	2.74	2.41	1.82	1.76	1.95	1.97	1.74	1.59
X	D	15	6.56	4.33	4.12	4.51	4.73	4.11	3.03	2.26	1.88	1.74	1.84	1.90	1.68	1.66
X	E	15	7.22	7.81	5.86	5.65	6.32	5.47	5.00	2.01	2.49	2.03	1.81	2.02	1.95	2.99
Y	A	15	6.02	7.85	5.71	5.33	5.90	5.30	5.70	6.29	7.78	4.92	4.49	5.55	4.49	3.39
Y	B	15	6.94	4.30	4.27	4.43	4.80	4.41	3.33	6.19	4.14	3.66	3.71	4.39	3.36	2.63
Y	C	15	6.90	3.75	4.08	4.38	4.73	4.14	2.72	6.60	3.56	3.55	3.80	4.32	3.32	2.54
Y	D	15	5.61	4.50	4.18	4.13	4.47	4.15	3.64	6.55	4.18	3.72	3.79	4.45	3.48	2.69
Y	E	15	6.33	7.37	5.59	5.23	5.83	5.22	5.07	6.34	7.41	4.89	4.35	5.48	4.55	3.26

FIGURE 3. Case DDe

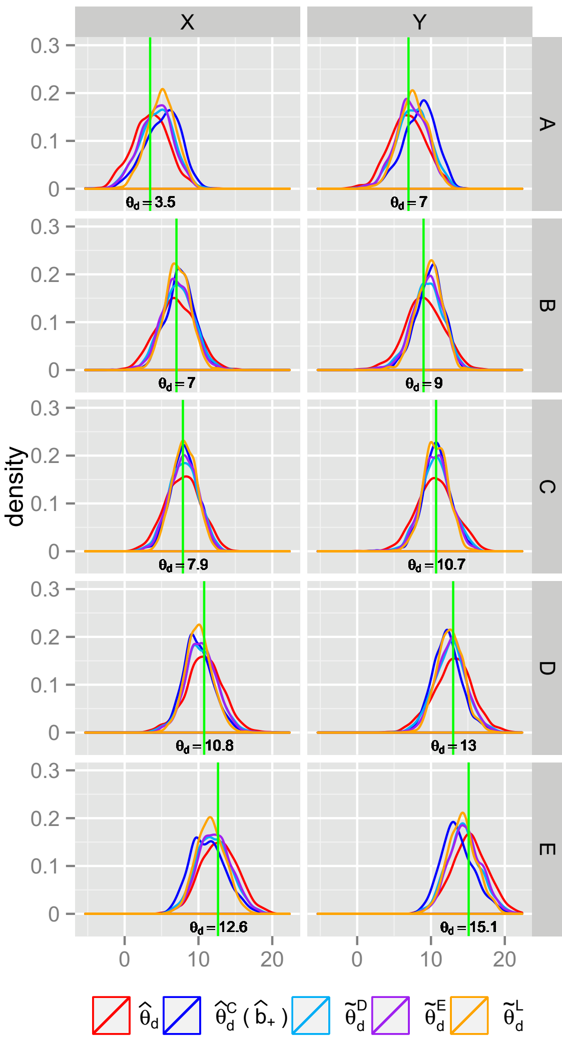

FIGURE 4. Case DDu

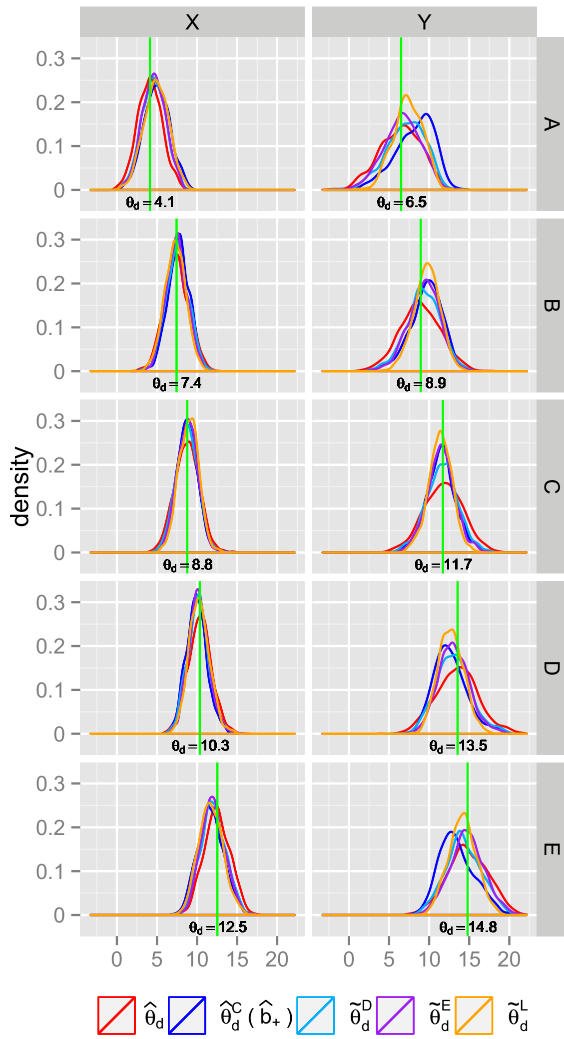

TABLE 2. Empirical MSE of Estimators for Case DD

			Case DDe							Case DDu						
		n_d	$\hat{\theta}_d$	$\hat{\theta}_d^C(\hat{b}_\dagger)$	$\tilde{\theta}_d^A$	$\tilde{\theta}_d^C$	$\tilde{\theta}_d^D$	$\tilde{\theta}_d^E$	$\tilde{\theta}_d^L$	$\hat{\theta}_d$	$\hat{\theta}_d^C(\hat{b}_\dagger)$	$\tilde{\theta}_d^A$	$\tilde{\theta}_d^C$	$\tilde{\theta}_d^D$	$\tilde{\theta}_d^E$	$\tilde{\theta}_d^L$
X	A	15	6.03	8.52	6.01	5.30	6.15	5.50	6.06	2.29	3.29	2.50	2.16	2.48	2.35	3.03
X	B	15	6.87	4.20	4.32	4.74	4.91	4.36	3.11	2.44	1.87	1.80	2.01	2.01	1.77	1.83
X	C	15	5.66	3.40	3.59	3.91	4.07	3.65	2.70	2.39	1.73	1.69	1.90	1.90	1.66	1.58
X	D	15	6.48	5.02	4.60	4.74	5.06	4.51	3.71	2.25	1.85	1.69	1.85	1.87	1.65	1.72
X	E	15	6.39	7.47	5.45	5.09	5.76	5.12	4.78	2.65	3.27	2.65	2.41	2.67	2.54	2.95
Y	A	15	6.18	7.15	5.11	4.90	5.57	4.75	4.19	6.94	8.82	5.71	5.17	6.26	5.39	4.09
Y	B	15	6.90	4.72	4.58	4.77	5.10	4.62	3.73	6.32	4.62	4.12	4.09	4.66	4.01	3.26
Y	C	15	6.65	3.64	3.82	4.15	4.47	3.85	2.79	5.86	3.40	3.35	3.44	3.99	3.08	2.26
Y	D	15	6.99	4.67	4.38	4.63	5.03	4.28	3.33	6.93	5.19	4.32	4.28	5.08	4.00	3.22
Y	E	15	6.07	7.29	5.29	5.01	5.56	4.99	4.71	6.37	6.63	4.80	4.45	5.34	4.52	3.37

FIGURE 5. Case SDe

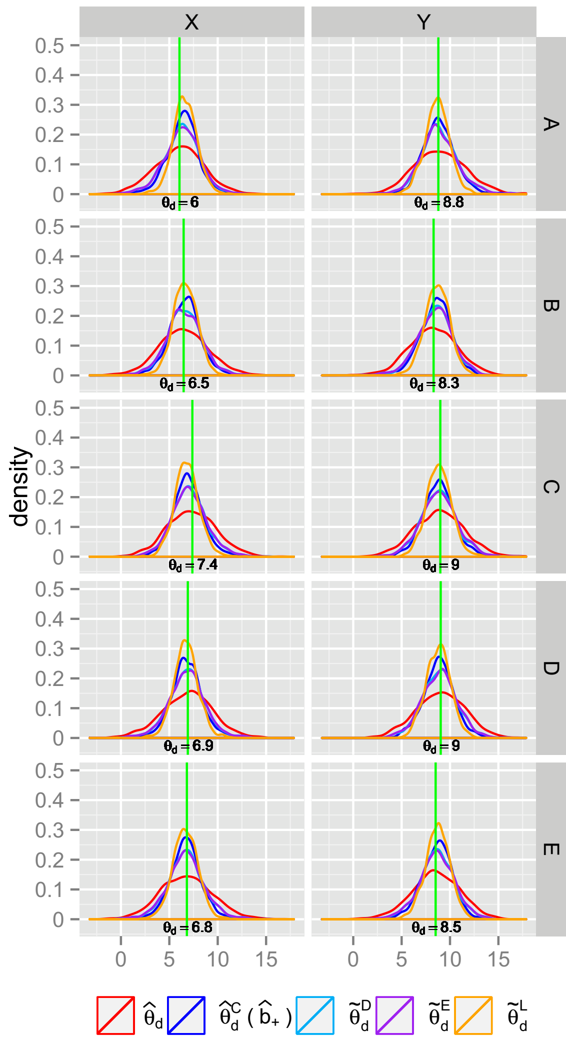

FIGURE 6. Case SDu

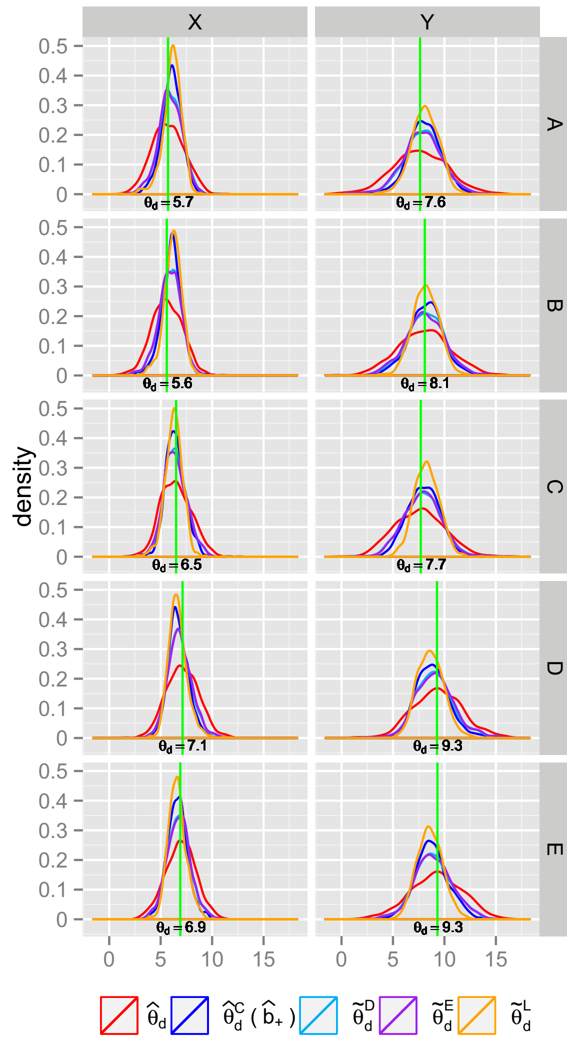

TABLE 3. Empirical MSE of Estimators for Case SD

			Case SDe							Case SDu						
		n_d	$\hat{\theta}_d$	$\hat{\theta}_d^C(\hat{b}_\dagger)$	$\tilde{\theta}_d^A$	$\tilde{\theta}_d^C$	$\tilde{\theta}_d^D$	$\tilde{\theta}_d^E$	$\tilde{\theta}_d^L$	$\hat{\theta}_d$	$\hat{\theta}_d^C(\hat{b}_\dagger)$	$\tilde{\theta}_d^A$	$\tilde{\theta}_d^C$	$\tilde{\theta}_d^D$	$\tilde{\theta}_d^E$	$\tilde{\theta}_d^L$
X	A	15	6.22	2.74	3.06	3.61	3.52	3.59	1.88	2.42	1.10	1.24	1.58	1.39	1.39	1.04
X	B	15	6.43	2.64	2.96	3.59	3.49	3.54	1.60	2.25	1.16	1.22	1.50	1.37	1.36	1.13
X	C	15	6.20	2.69	3.05	3.55	3.46	3.56	1.98	2.23	0.95	1.14	1.51	1.30	1.34	0.82
X	D	15	6.81	2.78	3.15	3.77	3.68	3.75	1.72	2.48	1.20	1.30	1.63	1.47	1.48	1.10
X	E	15	7.19	2.82	3.24	3.92	3.81	3.90	1.72	2.17	1.07	1.18	1.48	1.33	1.34	0.89
Y	A	15	7.09	2.89	3.24	3.66	3.79	3.87	1.66	7.04	3.11	3.48	3.42	3.95	4.00	2.06
Y	B	15	6.00	2.71	2.92	3.26	3.37	3.47	1.75	6.54	2.89	3.30	3.32	3.75	3.86	1.68
Y	C	15	6.58	2.78	3.22	3.63	3.69	3.77	1.72	5.98	2.79	3.07	3.10	3.52	3.48	1.85
Y	D	15	6.28	2.62	2.86	3.24	3.34	3.39	1.66	5.85	2.91	3.31	3.22	3.65	3.70	2.09
Y	E	15	6.12	2.57	2.90	3.27	3.40	3.50	1.68	6.39	2.86	3.24	3.22	3.66	3.63	2.08

5. Conclusion

We reviewed Longford's multivariate composite estimation method and modified the estimation of the covariance matrix $\boldsymbol{\Sigma}$, using the Ledoit-Wolf approach of shrinkage covariance matrices. The simulation study encompasses various data designs to show the performance of the baseline univariate method, the basic linear mixed-effects model, and multivariate estimators using several different target covariance matrices. The empirical MSEs calculated from this study suggest that the target covariance matrices with specified covariance entries perform slightly better than those with only diagonal variance entries. While the linear mixed-effects model yields the smallest errors overall, the multivariate estimators are less prone to incurred bias. The univariate shrinkage estimator is more efficient than the multivariate estimators only under the condition of across-district similarity. For other cases with across-district dissimilarity, the multivariate estimators incur much less bias, even in districts with extreme underlying parameters.

In data applications, multivariate estimators have an advantage of fewer data requirements. Data aggregated by district and subpopulation are sufficient for developing multivariate estimators. On the other hand, the linear mixed-effects model requires micro-level data as an input. Thus, while the linear mixed-effects model yields smaller errors, the multivariate composite estimation method has the advantage of data flexibility.

Future extensions to this study could involve the addition of auxiliary variables in the multivariate composite estimation method. Please see Appendix A for a review of Longford's auxiliary extension to the multivariate composite method [4]. This extension of the multivariate method could be compared to more complicated linear mixed-effects models that incorporate additional variables as random and fixed effects.

APPENDIX A. MULTIVARIATE COMPOSITE ESTIMATION WITH AUXILIARY VARIABLES

In this section, we review Longford's extension of the multivariate composite estimation method, using auxiliary variables [4]. The composite form with auxiliary variables is

$$\tilde{\theta}_d^A = \begin{bmatrix} \hat{\boldsymbol{\theta}}_d \\ \hat{\boldsymbol{\theta}}_d^{(a)} \end{bmatrix}^T \begin{pmatrix} \mathbf{w} - \mathbf{b}_d \\ \mathbf{0} - \mathbf{b}_d^{(a)} \end{pmatrix} + \begin{bmatrix} \hat{\boldsymbol{\theta}} \\ \hat{\boldsymbol{\theta}}^{(a)} \end{bmatrix}^T \begin{pmatrix} \mathbf{b}_d \\ \mathbf{b}_d^{(a)} \end{pmatrix} \tag{A.1}$$

where $\hat{\boldsymbol{\theta}}_d$ and $\hat{\boldsymbol{\theta}}_d^{(a)}$ are the direct and auxiliary estimators for district d and the nation, respectively. Note that this notation differentiating the direct and auxiliary estimators extends to other components in the augmented optimal weight vector.

By minimizing eMSE $\left\{\tilde{\theta}_d^A\left(\left[\mathbf{b}_d^T\ \mathbf{b}_d^{(a)T}\right]^T\right); \boldsymbol{\theta}_d^T\mathbf{w}\right\}$ with respect to the augmented weight vector $\left[\mathbf{b}_d^T\ \mathbf{b}_d^{(a)T}\right]^T$, we obtain the following optimal augmented weight vector, which is the auxiliary analog of (1.9)

$$\begin{bmatrix} \mathbf{b}_d^\star \\ \mathbf{b}_d^{(a)\star} \end{bmatrix} = \tilde{\mathbf{D}}_d^{-1}(\mathbf{I} - \tilde{\mathbf{N}}_d)\tilde{\mathbf{V}}_d \begin{bmatrix} \mathbf{w} \\ \mathbf{0} \end{bmatrix} \tag{A.2}$$

where

$$\tilde{\mathbf{D}}_d = \tilde{\mathbf{V}}_d + \tilde{\boldsymbol{\Sigma}} + \tilde{\mathbf{V}} - \tilde{\mathbf{N}}_d\tilde{\mathbf{V}}_d - \tilde{\mathbf{V}}_d\tilde{\mathbf{N}}_d \tag{A.3}$$

and

$$\tilde{\mathbf{V}}_d = \begin{bmatrix} \mathbf{V}_d & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_d^{(a)} \end{bmatrix} \qquad \tilde{\mathbf{V}} = \begin{bmatrix} \mathrm{Var}(\hat{\boldsymbol{\theta}}) & \mathbf{0} \\ \mathbf{0} & \mathrm{Var}(\hat{\boldsymbol{\theta}}^{(a)}) \end{bmatrix}$$

$$\tilde{\boldsymbol{\Sigma}} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}^{(0,a)} \\ \boldsymbol{\Sigma}^{(a,0)} & \boldsymbol{\Sigma}^{(a)} \end{bmatrix} \qquad \tilde{\mathbf{N}}_d\tilde{\mathbf{V}}_d = \begin{bmatrix} \mathbf{N}_d\mathbf{V}_d & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_d^{(a)}\mathbf{V}_d^{(a)} \end{bmatrix}$$

Thus, we can rewrite (A.3) as

$$\tilde{\mathbf{D}}_d = \begin{bmatrix} \mathbf{D}_d & \boldsymbol{\Sigma}^{(0,a)} \\ \boldsymbol{\Sigma}^{(a,0)} & \mathbf{D}_d^{(a)} \end{bmatrix} \tag{A.4}$$

where $\boldsymbol{\Sigma}^{(0,a)} = (\boldsymbol{\Sigma}^{(a,0)})^T$ is the covariance submatrix of $\tilde{\boldsymbol{\Sigma}}$ for the direct estimator $\hat{\boldsymbol{\theta}}_d$ and the auxiliary variables $\hat{\boldsymbol{\theta}}_d^{(a)}$.

Since $\mathbf{b}_d^{(a)\star} = \mathbf{0}$, we can focus on the nonzero component of (A.2), which is

$$\mathbf{b}_d^\star = \mathbf{F}_d(\mathbf{I} - \mathbf{N}_d)\mathbf{V}_d\mathbf{w} \tag{A.5}$$

where

$$\mathbf{F}_d = \left(\mathbf{D}_d - \boldsymbol{\Sigma}^{(0,a)}\left(\mathbf{D}_d^{(a)}\right)^{-1}\boldsymbol{\Sigma}^{(a,0)}\right)^{-1} \tag{A.6}$$

After substituting (A.5) into the expected mean squared error function and extracting $\mathbf{w}$, we obtain the analog of (1.18)

$$\tilde{\boldsymbol{\theta}}_d = [\mathbf{I} - \mathbf{V}_d(\mathbf{I} - \mathbf{N}_d)\mathbf{F}_d]\hat{\boldsymbol{\theta}}_d + \mathbf{V}_d(\mathbf{I} - \mathbf{N}_d)\mathbf{F}_d\hat{\boldsymbol{\theta}} \tag{A.7}$$

In estimating $\tilde{\boldsymbol{\Sigma}}$, one might adopt the Ledoit-Wolf approach of shrinkage covariance matrices [3]. As we have shown in the simulation study in Section 4, the target covariance matrices with specified covariance entries tend to yield smaller errors than the target matrices with only diagonal variance entries. Given that the auxiliary extension involves potentially highly correlated direct and auxiliary data, target covariance matrices with specified covariance entries could be advantageous.

References

[1] Bates, D., Maechler, M., Bolker, B., and Walker, S. *lme4: Linear mixed-effects models using Eigen and S4*, 2013. R package version 1.0-5.

[2] Hogue, M., and Li, D. Composite Shrinkage Estimation: Review of Theory with Simulation Study and Empirical Applications. Economic and Statistical Working Paper Series. Bureau of Economic and Business Research, University of Utah, 2014.

[3] Ledoit, O., and Wolf, M. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance 10* (2003), 603–621.

[4] Longford, N. T. Multivariate shrinkage estimation of small area means and proportions. *Journal of the Royal Statistical Society. Series A (Statistics in Society) 162*, 2 (1999), 227–245.

[5] Longford, N. T. *Missing Data and Small-Area Estimation: Modern Analytical Equipment for the Survey Statistician.* Springer, New York, 2005.

[6] Schäfer, J., and Strimmer, K. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genetics. *Statistical Applications in Genetics and Molecular Biology 4* (2005), 1–30.

[7] Wu, L. *Mixed Effects Models for Complex Data.* Chapman & Hall/CRC Press, Boca Raton, 2009.