SMALL AREA MEAN ESTIMATION BASED ON A NESTED-ERROR REGRESSION MODEL

Paddison C. F. Wong, Statistics Canada, RHC-16K, 120 Parkdale Avenue, Ottawa, Ontario, Canada K1A 0T6

Kevin J. Keen, Department of Epidemiology and Biostatistics, Case Western Reserve University, Rammelkamp Center for Education and Research, MetroHealth Medical Center, 2500 MetroHealth Drive, Cleveland, Ohio 44109-1998

Key Words: Maximum likelihood estimator, mixed linear model, nested-error regression model, random intercept model

INTRODUCTION

Attempts at small-area estimation by direct survey estimators are not successful because small sample sizes yield considerably large standard errors. The development of efficient small-area estimation methods is a challenging statistical problem. One way to approach this problem is to attempt to increase precision by borrowing information across similar small areas. However, some of these estimators can be heavily biased in the design-based framework for some of the areas. There is a demand for unbiased estimators with better efficiency and a willingness to use auxiliary information to satisfy this demand.

In the general mixed linear model context, small-area means can be efficiently estimated by best linear unbiased predictors if variance-component parameters are known. If not, as is typically the case, they can be estimated and replace the parameters in the predictors.

A brief review on the nested-error regression model and unbiased estimation of the small-area means is given in the next section followed by the development of a weighted unbiased estimator of the group-effect variance component for the random intercept model after identifying the relationship between the random intercept and nested-error regression models. To improve the efficiency of estimating the small-area means, a two-stage small-area estimator that is a linear combination of two small-area estimators with different weights for the estimator of the group effect is presented and a formula for its mean squared error approximation is also given. A comparison in terms of the relative efficiencies of different small-area mean estimators to the maximum likelihood estimator with Monte Carlo simulations is then presented.

THE NESTED-ERROR REGRESSION MODEL

Consider the nested-error regression model with a single random effect, ν_i, for the ith small area, or small group:

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \nu_i + \epsilon_{ij}, \quad i = 1, \ldots, t, \quad j = 1, \ldots, n_i, \tag{2.1}$$

where y_{ij} is the observed response for the jth sampled unit in the ith group, $\mathbf{x}_{ij} = (1, x_{ij1}, \cdots, x_{ijk})'$ is a $(k+1)$-vector of corresponding covariates, $\boldsymbol{\beta} = (\alpha, \beta_1, \cdots, \beta_k)'$ is a $(k+1)$-vector of unknown regression coefficients, and n_i is the size of the ith group. The group effects $\{\nu_i\}$ are identically distributed with mean 0 and variance σ_ν^2, the random errors $\{\epsilon_{ij}\}$ are identically distributed with mean 0 and variance σ_ϵ^2 with $\{\nu_i\}$ and $\{\epsilon_{ij}\}$ mutually independent. For convenience, define matrices $\mathbf{X}' = (\mathbf{x}_{11}, \ldots, \mathbf{x}_{tn_t})$ and $\mathbf{Z} = \mathrm{diag}(\mathbf{1}_{n_1}, \ldots, \mathbf{1}_{n_t})$ where $\mathbf{1}_{n_i}$ is a n_i-vector with a value of 1 for each element. If one lets $\mathbf{y} = (y_{11}, \ldots, y_{tn_t})'$, then the model in (2.1) can be expressed in matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\nu} + \boldsymbol{\epsilon} \tag{2.2}$$

with $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_t)'$ and $\boldsymbol{\epsilon} = (\epsilon_{11}, \ldots, \epsilon_{tn_t})'$ which are independently distributed with mean vector $\mathbf{0}$ and covariance matrices $\mathbf{G} = \sigma_\nu^2\mathbf{I}_t$ and $\mathbf{R} = \sigma_\epsilon^2\mathbf{I}_N$, respectively. The total number of observations in the sample is $N = \sum_{i=1}^t n_i$. For known $\mathbf{G}$ and $\mathbf{R}$, Henderson (1975) showed that the best linear unbiased estimator, or predictor, (BLUE, or BLUP) of $\boldsymbol{\mu} = \mathbf{l}'\boldsymbol{\beta} + \mathbf{m}'\boldsymbol{\nu}$ is given by

$$\hat{\boldsymbol{\mu}} = \mathbf{l}'\tilde{\boldsymbol{\beta}} + \mathbf{m}'\mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}\right) \tag{2.3}$$

where $\tilde{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ is the generalized least squares estimator of $\boldsymbol{\beta}$. Let $\overline{\mathbf{X}}_i{}'$ be the vector of known means of $\mathbf{x}_{ij}$ for the ith group and let the sample mean vector $\overline{\mathbf{x}}_i = n_i^{-1}\sum_{j=1}^{n_i}\mathbf{x}_{ij}$. Define $\boldsymbol{\mu}$ to be the vector of small-area means in the population with component

$$\mu_i = \overline{\mathbf{X}}_i'\boldsymbol{\beta} + \nu_i \tag{2.4}$$

for the ith area. An unbiased estimator of μ_i is given by

$$\overline{\mathbf{X}}_i'\tilde{\boldsymbol{\beta}} + \mathbf{m}_i'\mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}\right) \tag{2.5}$$

where $\mathbf{m}_i$ is a t-vector of zero with the ith element equal to 1. Note that $\tilde{\boldsymbol{\beta}}$ is always a function of the unknown parameters σ_ν^2 and σ_ϵ^2. It is natural to replace them with corresponding unbiased estimators.

ESTIMATION OF VARIANCE COMPONENTS

The method of fitting constants was used in Prasad and Rao (1990) for the estimation of σ_ϵ^2 and σ_ν^2. Their estimators are given by

$$\tilde{\sigma}_\epsilon^2 = \frac{\mathbf{y}'\left[\mathbf{I} - \mathbf{M}(\mathbf{M}'\mathbf{M})^{-}\mathbf{M}'\right]\mathbf{y}}{N - \mathrm{r}(\mathbf{M})}\ , \tag{3.1}$$

and

$$\tilde{\sigma}_\nu^2 = \frac{\sum_{i=1}^{t}\sum_{j=1}^{n_i}\hat{u}_{ij}^2 - (n-k)\tilde{\sigma}_\epsilon^2}{n - \mathrm{tr}\left[(\mathbf{X}'\mathbf{X})^{-}\sum_{i=1}^{t} n_i^2 \overline{\mathbf{x}}_i\ \overline{\mathbf{x}}_i'\right]} \tag{3.2}$$

where $\mathbf{M} = (\mathbf{X} \mid \mathbf{Z})$, $\mathrm{r}(\mathbf{M})$ is the rank of $\mathbf{M}$ and $\sum_{i=1}^{t}\sum_{j=1}^{n_i}\hat{u}_{ij}^2 = \mathbf{y}'\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\right]\mathbf{y}$.

With $\alpha_i = \alpha + \nu_i$, $\{\alpha_i\}$ can be considered as random intercepts identically distributed with mean α and variance σ_ν^2. Upon setting $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_t)'$ and $\boldsymbol{\beta}_o = (\beta_1, \ldots, \beta_k)'$, the model in (2.2) can be written as a random intercept model as in

$$\mathbf{y} = \mathbf{X}_\alpha\boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{3.3}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}_o', \boldsymbol{\alpha}')'$ is a $(k+t)$-vector with $\mathbf{X}_\alpha$ is a corresponding design matrix of covariates that satisfies (3.3). Let $\mathbf{D}_\alpha = (\mathbf{0}_{t\times k} \mid \mathbf{I}_t)$ with $\mathbf{0}_{t\times k}$ a matrix of zeros and $\mathbf{I}_t$ a $(t \times t)$ identity matrix so that

$$\mathrm{E}\left[\mathbf{y}\right] = \mathbf{X}\boldsymbol{\beta} = \mathbf{X}_\alpha\boldsymbol{\theta}_o$$

where $\boldsymbol{\theta}_o = (\boldsymbol{\beta}_o', \alpha\mathbf{1}')'$. An unbiased weighted estimator of σ_ν^2 based on model (3.3) is given by

$$s_\nu^2 = \left[(t-1)\left(\overline{\omega} - \frac{s_\omega^2}{t\,\overline{\omega}}\right)\right]^{-1}\left\{\sum_{i=1}^{t}\omega_i(\hat{\alpha}_i - \hat{\alpha})^2 - s_\epsilon^2\mathrm{tr}\Big[\left(\boldsymbol{\Omega} - N_\omega^{-1}\boldsymbol{\omega}\boldsymbol{\omega}'\right)\mathbf{D}_\alpha(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\mathbf{D}_\alpha'\Big]\right\} \tag{3.4}$$

with

$$s_\epsilon^2 = \frac{\mathbf{y}'\left[\mathbf{I} - \mathbf{X}_\alpha\left(\mathbf{X}_\alpha'\mathbf{X}_\alpha\right)^{-1}\mathbf{X}_\alpha'\right]\mathbf{y}}{N - \mathrm{r}\left(\mathbf{X}_\alpha\right)} \tag{3.5}$$

where ω_i is the weight associated for the ith group, $\boldsymbol{\omega}' = (\omega_1, \omega_2, \ldots \omega_t)$, $\boldsymbol{\Omega} = \mathrm{diag}(\omega_1, \omega_2, \ldots \omega_t)$, $N_\omega = \sum_{i=1}^{t}\omega_i$, $\overline{\omega} = t^{-1}\sum_{i=1}^{t}\omega_i$, $s_\omega^2 = (t-1)^{-1}\sum_{i=1}^{t}(\omega_i - \overline{\omega})^2$, and $\hat{\alpha} = N_\omega^{-1}\sum_{i=1}^{t}\omega_i\hat{\alpha}_i$, where $\hat{\alpha}_i$ is the ordinary least squares estimator of α_i for the ith area and $\mathrm{r}\left(\mathbf{X}_\alpha\right)$ is the rank of $\mathbf{X}_\alpha$. Both s_ν^2 and s_ϵ^2 are unbiased for σ_ν^2 and σ_ϵ^2, respectively.

It is possible for $\tilde{\sigma}_\nu^2$ and s_ν^2 to be negative, therefore, it is common practice to take instead $\max(0, \tilde{\sigma}_\nu^2)$ and $\max(0, s_\nu^2)$ as estimators of σ_ν^2. Nevertheless, as sample sizes increase, both $\tilde{\sigma}_\nu^2$ and s_ν^2 tend to be non-negative almost surely.

ESTIMATION OF SMALL-AREA MEANS

To estimate the small-area means using equation (2.5), Prasad and Rao (1990) replaced σ_ϵ^2 and σ_ν^2 in $\mathbf{G}, \mathbf{V}$ and $\tilde{\boldsymbol{\beta}}$ with their estimators $\tilde{\sigma}_\epsilon^2$ and $\tilde{\sigma}_\nu^2$ in (3.1) and (3.2). Under the random intercept model (3.3), the estimator s_ν^2 in equation (3.4) has an advantage in that the weights $\{\omega_i\}$ can be modified to adjust for differences among the small areas or equivalently the effects of the random intercept estimators $\{\hat{\alpha}_i\}$. The efficiency of small-area mean estimation can be further improved with a convex combination of two different estimators of σ_ν^2, namely $s_{\nu 1}^2$ and $s_{\nu 2}^2$ with different weights, $\{\omega_i^{(1)}\}$ and $\{\omega_i^{(2)}\}$, respectively, as in equation (3.4). An estimator of the mean of the ith small area by this approach is

$$\tilde{\mu}_i^c = (1 - a_i)\,\tilde{\mu}_{i1} + a_i\,\tilde{\mu}_{i2} \tag{4.1}$$

where $\tilde{\mu}_{i1}$ and $\tilde{\mu}_{i2}$ are estimators of μ_i using $s_{\nu 1}^2$ and $s_{\nu 2}^2$ respectively. Obviously the choice of a_i would affect the value of the mean squared error of $\tilde{\mu}_i^c$. An efficient estimator of μ_i is the one with the smallest mean squared error. The minimum mean squared error of $\tilde{\mu}_i^c$

$$\mathrm{MSE}[\tilde{\mu}_i^c] = \Big(\mathrm{MSE}[\tilde{\mu}_{i1}]\,\mathrm{MSE}[\tilde{\mu}_{i2}] - \Big\{\mathrm{E}\left[(\tilde{\mu}_{i1} - \mu_i)(\tilde{\mu}_{i2} - \mu_i)\right]\Big\}^2\Big) \Big/ \Big\{\mathrm{MSE}[\tilde{\mu}_{i1}] + \mathrm{MSE}[\tilde{\mu}_{i2}] - 2\,\mathrm{E}\left[(\tilde{\mu}_{i1} - \mu_i)(\tilde{\mu}_{i2} - \mu_i)\right]\Big\} \tag{4.2}$$

Table 1. Average Relative efficiency of $\tilde{\mu}^c$ with respect to $\tilde{\mu}$ over 20 small areas with different distributions of random effects for ρ between 0.00 and 0.20 with known population mean $\{\overline{\mathbf{X}}_i\}$

ρ	**Norm.**	**Dbl.Exp.**	**Exp.**	**Unif.**
0.00	1.0636	1.1007	1.1287	1.0522
0.05	1.0186	1.0289	1.0368	1.0146
0.10	1.0094	1.0135	1.0170	1.0079
0.15	1.0052	1.0074	1.0086	1.0042
0.20	1.0023	1.0041	1.0049	1.0013

is attained when

$$a_i = \Big\{ \text{MSE}[\tilde{\mu}_{i1}] - \text{E}\Big[(\tilde{\mu}_{i1} - \mu_i)(\tilde{\mu}_{i2} - \mu_i)\Big]\Big\} \Big/ \Big\{ \text{MSE}[\tilde{\mu}_{i1}] + \text{MSE}[\tilde{\mu}_{i2}] - 2\,\text{E}\Big[(\tilde{\mu}_{i1} - \mu_i)(\tilde{\mu}_{i2} - \mu_i)\Big]\Big\} \,. \tag{4.3}$$

RESULTS OF MONTE CARLO STUDY

A Monte Carlo simulation that mimics the data set of Battese et al. (1988) under the nested-error regression model with a single covariate, $y_{ij} = \alpha + \beta x_{ij} + \nu_i + \epsilon_{ij}$, was conducted to study the relative efficiency of the two-stage estimator combining $s^2_{\nu 1}$ and $s^2_{\nu 2}$ with respect to the one with $\tilde{\sigma}^2_\nu$. Let y_{ij} equal acreage of corn cultivation and x_{ij} equal the number of pixels of corn cultivation for the jth unit of county i from satellite imagery. The three counties with only 1 unit were pooled together as done in Prasad and Rao (1990). The number of small areas t was increased to 20 from 10 by duplication as also done by Prasad and Rao (1990). The value of the intercept α and the slope β were set to the estimates obtained by Battese et al. (1988), that is, 5.5 and 0.388, respectively. Ten thousand independent sets of ν_i and ϵ_{ij} with zero means and different intraclass correlation $\rho = \sigma^2_\nu/(\sigma^2_\nu + \sigma^2_\epsilon)$ from 0.00 to 0.95 were generated to obtain 10,000 sets of y_{ij} using the covariate values of Battese et al. (1988) for each of 4 distributional assumptions: normal, double-exponential, exponential and uniform. The weights $\{\omega^{(1)}_i\}$ for the estimator $s^2_{\nu 1}$ in $\tilde{\mu}_{i1}$ are equal to 1. This is the uniform weighting scheme which is thought to be best for a large intraclass correlation ρ and shown to be such in Keen (1996) in a similar situation but without covariates. The weights $\{\omega^{(2)}_i\}$ for the estimator $s^2_{\nu 2}$ in $\tilde{\mu}_{i2}$ are equal to $n_i(n_i - 1)$ in a pairwise weighting scheme which is thought to be best for a small intraclass correlation ρ based on the simulations in Keen (1996) without covariates. This latter scheme implicitly deletes single-member families.

From equation (2.5), one obtains the two-stage estimators of the small-area means, $\tilde{\boldsymbol{\mu}}^c$ with equation (4.1) and $\tilde{\boldsymbol{\mu}}$ by the approach of Prasad and Rao (1990). These estimators involve the population means of the covariates $\{\mathbf{x}_{ij}\}$ which are assumed to be known or available from censuses. The average relative efficiencies are summarized in Table 1 based on different distributions for $0.00 \le \rho \le 0.20$. The proposed estimator $\tilde{\boldsymbol{\mu}}^c$ is more efficient than $\tilde{\boldsymbol{\mu}}$ as the intraclass correlation decreases. The average relative efficiency in favor of the proposed estimator over 20 small areas can be as high as 112.87% with $\{\overline{\mathbf{X}}_i\}$ for the exponential distribution at $\rho = 0.00$. Similar pattern of efficiency can be found for the other distributions considered.

Figure 1 summarizes the comparison of $\tilde{\boldsymbol{\mu}}^c$ to $\tilde{\boldsymbol{\mu}}$ when $\{\overline{\mathbf{X}}_i\}$ are used for the full range of ρ. The estimator $\tilde{\boldsymbol{\mu}}^c$ is more efficient than $\tilde{\boldsymbol{\mu}}$ for small intraclass correlations. For other values of ρ, it is equally efficient as $\tilde{\boldsymbol{\mu}}$. This is not only true for the normal distribution, it is also true for other 3 distributions under studied. For some small areas, the relative efficiency can be as high as 128% for the exponential distribution and 140% for the double-exponential distribution. Evidently, $\tilde{\boldsymbol{\mu}}^c$ has a better efficiency profile than the existing estimator $\tilde{\boldsymbol{\mu}}$.

Let $\hat{\boldsymbol{\mu}}^{Vm}$ be the BLUP in which the unknown variance components are replaced by their maximum likelihood estimators (MLE's). Figure 2 is the comparison of $\tilde{\boldsymbol{\mu}}^c$ and $\tilde{\boldsymbol{\mu}}$ with $\hat{\boldsymbol{\mu}}^{Vm}$ in the case of normality. Note that $\tilde{\boldsymbol{\mu}}^c$ is more efficient than the MLE for all values of the intraclass correlation whereas $\tilde{\boldsymbol{\mu}}$ is only more efficient than the MLE when $\rho \ge 0.20$.

CONCLUSIONS AND FUTURE DIRECTIONS

The convex combination, which is implicitly a shrinkage approach, has more flexibility in adjusting for the impact of random effects on the small areas by adaptively balancing two different weighting schemes. The new results from the Monte Carlo simulations show that the proposed estimator is asymptotically more efficient than the maximum likelihood estimator assuming a normal distribution. More importantly, no matter whether the population means of the covariates are known or not, the proposed estimator $\tilde{\boldsymbol{\mu}}^c$ is more efficient than $\tilde{\boldsymbol{\mu}}$ of Prasad and Rao (1990) when the intraclass correlation ρ is less than 0.20 for all four distributions studied in terms of their mean squared errors. Despite whether $\{\nu_i\}$ or $\{\epsilon_{ij}\}$ has more impact on the response, $\tilde{\boldsymbol{\mu}}^c$ should be the choice among the estimators studied for small-area estimation.

We further propose to extend the application of this new technique for estimating the small-area means by convex combination of two unbiased estimators to other mixed models, including the nonlinear models. Results obtained by using the sample counterparts $\overline{\mathbf{x}}_{ij}$, when the population means of the covariates $\overline{\mathbf{X}}_{ij}$ are unknown, will also be studied.

ACKNOWLEDGEMENTS

The authors would like to thank Statistics Canada for the support of this research. This work was also supported in part by a U.S. Public Health Service resource grant P41 RR03655 from the National Center of Research Resources, research grants GM 28356 and GM 31575 from the National Institute of General Medical Sciences and a research grant from the National Sciences and Engineering Research Council of Canada.

APPENDIX: TECHNICAL DETAILS

A.1 Proof of the Unbiasedness of s_ϵ^2

Note that the estimator of σ_ϵ^2 in equation (3.5) can be written as

$$s_\epsilon^2 = \mathbf{y}'\mathbf{A}\mathbf{y}$$

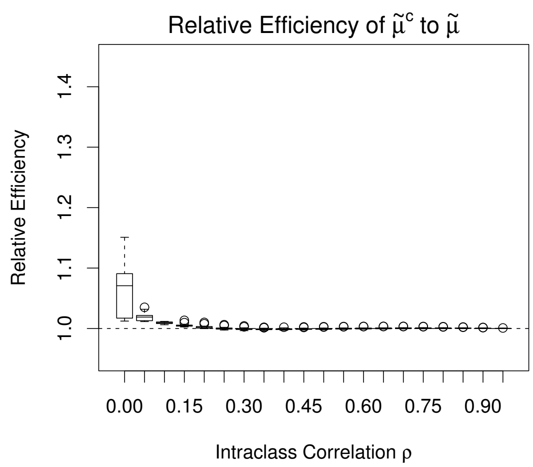
where

$$\mathbf{A} = \frac{\mathbf{I} - \mathbf{X}_\alpha \left(\mathbf{X}_\alpha'\mathbf{X}_\alpha\right)^{-1} \mathbf{X}_\alpha'}{N - \mathrm{r}\left(\mathbf{X}_\alpha\right)} \,. \tag{A.1}$$

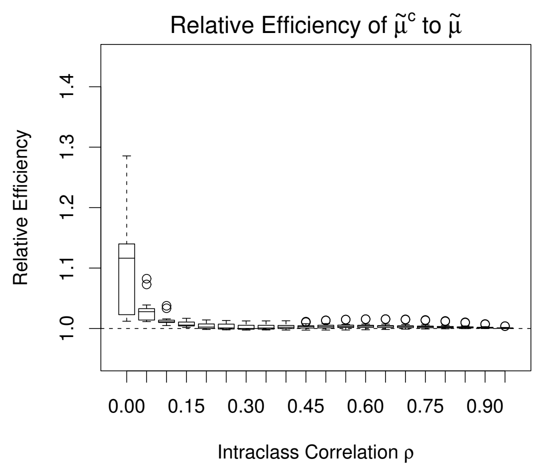
The expectation of s_ϵ^2 is then given by

$$\mathrm{E}\left[s_\epsilon^2\right] = \mathrm{tr}[\mathbf{AV}] + \boldsymbol{\theta}_o'\mathbf{X}_\alpha'\mathbf{A}\mathbf{X}_\alpha\boldsymbol{\theta}_o \tag{A.2}$$

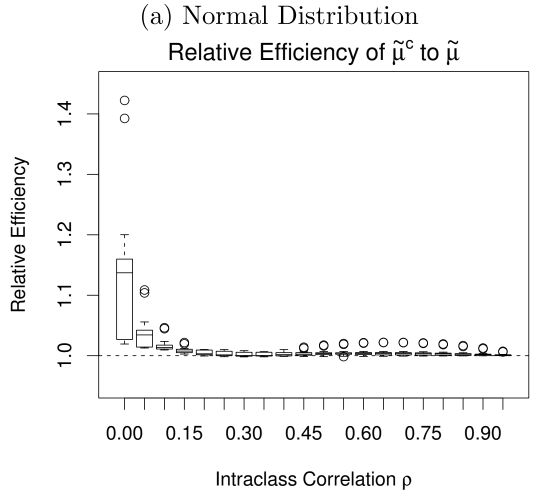
where $\mathbf{V} = \mathrm{Var}\left[\mathbf{y}\right] = \sigma_\epsilon^2\mathbf{I} + \sigma_\nu^2\mathbf{X}_\alpha\mathbf{D}_\alpha'\mathbf{D}_\alpha\mathbf{X}_\alpha'$ and $\boldsymbol{\theta}_o = (\boldsymbol{\beta}_o', \alpha\mathbf{1}')'$. Note that $\boldsymbol{\theta}_o'\mathbf{X}_\alpha'\mathbf{A}\mathbf{X}_\alpha\boldsymbol{\theta}_o = 0$ and it is easy to verify $\mathrm{tr}[\mathbf{AV}] = \sigma_\epsilon^2$.



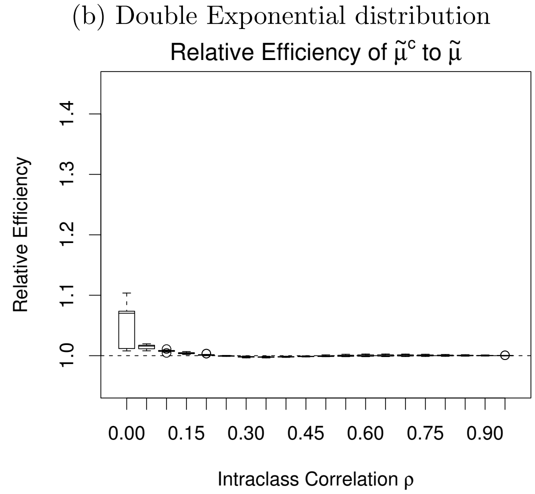
(a) Normal Distribution



(b) Double Exponential distribution

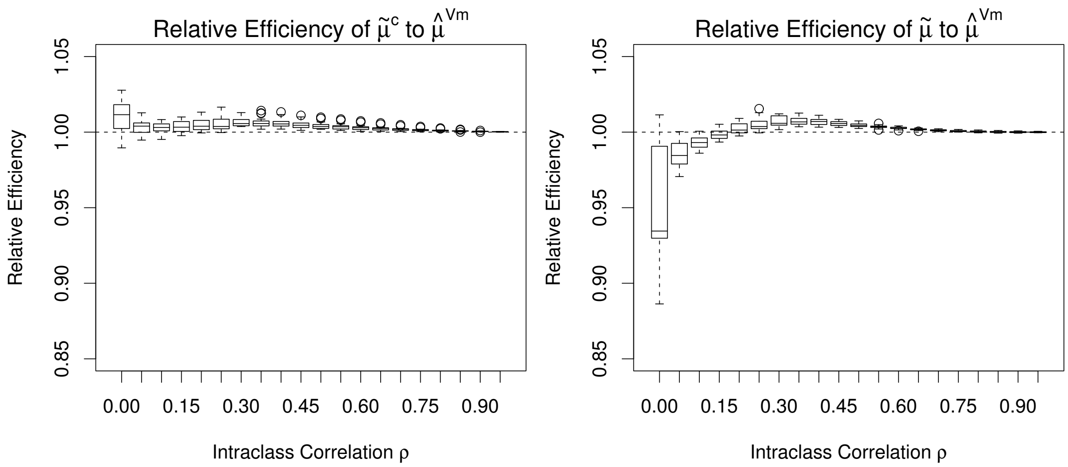


(c) Exponential Distribution



(d) Uniform Distribution

Figure 1. Boxplots of relative efficiencies of $\tilde{\boldsymbol{\mu}}^c$ with respect to $\tilde{\boldsymbol{\mu}}$ over 20 areas for intraclass correlation coefficient from 0.00 to 0.95 with 4 distributions of random effects assuming the population mean of the covariates in each area is known.



(a) $\tilde{\mu}^c$ with respect to $\hat{\mu}^{Vm}$ (b) $\tilde{\mu}$ with respect to $\hat{\mu}^{Vm}$

Figure 2. Boxplots of relative efficiencies of a) $\tilde{\boldsymbol{\mu}}^c$ with respect to $\hat{\boldsymbol{\mu}}^{Vm}$ and b) $\tilde{\boldsymbol{\mu}}$ with respect to $\hat{\boldsymbol{\mu}}^{Vm}$ over 20 areas for intraclass correlation coefficient from 0.00 to 0.95 with normal distribution of random effects assuming the population mean of the covariates in each area is known.

A.2 Proof of the Unbiasedness of s_ν^2

For convenience, write $\boldsymbol{\Omega}_* = \left(\boldsymbol{\Omega} - N_\omega^{-1}\boldsymbol{\omega}\boldsymbol{\omega}'\right)$ and $\mathbf{D}_* = \mathbf{D}_\alpha(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}$. Rewrite $\sum_{i=1}^t \omega_i(\hat{\alpha}_i - \hat{\alpha})^2 = \mathbf{y}'\mathbf{C}\mathbf{y}$ where $\mathbf{C} = \mathbf{X}_\alpha\mathbf{D}_*'\boldsymbol{\Omega}_*\mathbf{D}_*\mathbf{X}_\alpha'$. With the expression of s_ν^2 in equation (3.4), the expectation of s_ν^2 is given by

$$\mathrm{E}\left[s_\nu^2\right] = \frac{\mathrm{E}\left[\mathbf{y}'\mathbf{C}\mathbf{y}\right] - \sigma_\epsilon^2\,\mathrm{tr}\left[\boldsymbol{\Omega}_*\mathbf{D}_\alpha\mathbf{D}_*'\right]}{(t-1)\left[\overline{\omega} - \frac{s_w^2}{t\,\overline{\omega}}\right]} \tag{A.3}$$

because s_ϵ^2 is an unbiased estimator of σ_ϵ^2. The expectation of $\mathbf{y}'\mathbf{C}\mathbf{y}$ is given by

$$\begin{aligned}\mathrm{E}\left[\mathbf{y}'\mathbf{C}\mathbf{y}\right] &= \boldsymbol{\theta}_o{}'\mathbf{X}_\alpha'\mathbf{C}\mathbf{X}_\alpha\boldsymbol{\theta}_o + \mathrm{tr}\left[\mathbf{C}\mathbf{V}\right]\\ &= \sigma_\epsilon^2\,\mathrm{tr}\left[\boldsymbol{\Omega}_*\mathbf{D}_\alpha\mathbf{D}_*'\right]\\ &\quad + \sigma_\nu^2\,\mathrm{tr}\left[\mathbf{C}\mathbf{X}_\alpha\mathbf{D}_\alpha'\mathbf{D}_\alpha\mathbf{X}_\alpha'\right]\end{aligned} \tag{A.4}$$

because $\boldsymbol{\theta}_o'\mathbf{X}_\alpha'\mathbf{C}\mathbf{X}_\alpha\boldsymbol{\theta}$ is equal to 0. Note that $\sigma_\nu^2\,\mathrm{tr}\left[\mathbf{C}\mathbf{X}_\alpha\mathbf{D}_\alpha'\mathbf{D}_\alpha\mathbf{X}_\alpha'\right]$ can be written as $\sigma_\nu^2(t-1)\left[\overline{\omega} - (s_\omega^2/t\,\overline{\omega})\right]$ and therefore s_ν^2 is unbiased.

A.3 Optimal Choice of a_i

Let $\tilde{\mu}_i^c = (1-a_i)\,\tilde{\mu}_{i1} + a_i\,\tilde{\mu}_{i2}$ and the mean squared error can be written as

$$\begin{aligned}\mathrm{MSE}\;\left[\tilde{\mu}_i^c\right] &= \mathrm{E}\left[\left((1-a_i)\,\tilde{\mu}_{i1} + a_i\,\tilde{\mu}_{i2} - \mu_i\right)^2\right]\\ &= 2a_i(1-a_i)\,\mathrm{E}\left[(\tilde{\mu}_{i1}-\mu_i)(\tilde{\mu}_{i2}-\mu_i)\right]\\ &\quad + (1-a_i)^2\,\mathrm{MSE}\left[\tilde{\mu}_{i1}\right] + a_i^2\,\mathrm{MSE}\left[\tilde{\mu}_{i2}\right].\end{aligned} \tag{A.5}$$

Set the derivative of $\mathrm{MSE}[\tilde{\mu}_i^c]$ with respect to a_i to zero and it can be shown that the $\mathrm{MSE}[\tilde{\mu}_i^c]$ is at its minimum with

$$\begin{aligned}a_i =& \left\{\mathrm{MSE}[\tilde{\mu}_{i1}] - \mathrm{E}\left[(\tilde{\mu}_{i1}-\mu_i)(\tilde{\mu}_{i2}-\mu_i)\right]\right\}\\ &\Big/\Big\{\mathrm{MSE}[\tilde{\mu}_{i1}] + \mathrm{MSE}[\tilde{\mu}_{i2}]\\ &\quad - 2\,\mathrm{E}\left[(\tilde{\mu}_{i1}-\mu_i)(\tilde{\mu}_{i2}-\mu_i)\right]\Big\}.\end{aligned} \tag{A.6}$$

REFERENCES

Battese, G. E. and Fuller, W. A. (1988), "An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data," *Journal of the American Statistical Association*, **83**, 28–36.

Craig, A. T. (1943), "A Note on the Independence of Certain Quadratic Forms," *The Annals of Mathematical Statistics*, **14**, 195–197.

Ghosh, M., and Rao, J. N. K. (1994), "Small Area Estimation: An Appraisal," *Statistical Science*, **9**, 55–93.

Graybill, F. A. (1969), *Introduction to Matrices With Applications in Statistics*, Belmont, CA: Wadsworth.

Henderson, C. R. (1953), "Estimation of Variance and Covariance Components," *Biometrics*, **9**, 226–252.

Henderson, C. R. (1975), "Best Linear Unbiased Estimation and Prediction under a Selection Model," *Biometrics*, **31**, 423–447.

Keen, K. J. (1996), "Limiting the Effects of Single-Member Families in the Estimation of the Intraclass Correlation," *Biometrics*, **52**, 823–832.

Moura, F. A. S. and Holt, D. (1999), "Small Area Estimation Using Multilevel Models," *Survey Methodology*, **25**, 73–80.

Prasad, N. G. N. and Rao, J. N. K. (1990), "The Estimation of the Mean-Squared Error of Small-Area Estimators," *Journal of the American Statistical Association*, **85**, 163–171.

Searle, S. R. (1968), "Another Look at Henderson's Methods of Estimating Variance Components," *Biometrics*, **24**, 749–778.