# Small Area Estimation Under a Restriction

**Junyuan Wang, Merck Research Laboratories and Wayne A. Fuller, Iowa State University**
Junyuan Wang, Merck Research Laboratories, BL 3-2, PO Box 4,
West Point, PA 19486 (junyuan_wang@merck.com)

**Key Words:** Components-of-variance model; Best linear unbiased prediction; Benchmarking; Mixed linear models.

**Abstract:**

There are many situations in which it is desirable to derive reliable estimators for small geographical areas or small subpopulations, from existing survey data. The basic random effects model and corresponding small area predictors for small area estimation is introduced. There are also many situations in which it is necessary to have the total of the small area predictors equal to the total of the direct survey estimates for many small areas. This motivates the small area estimation under a restriction, which forces the sum of the small area predictors equal to certain benchmark. Several small area predictors under a restriction are reviewed. A criterion that unifies the derivation of these restricted predictors is proposed. The predictor that is the unique best linear unbiased estimator under the criterion is derived. The derivation of the mean square error (MSE) of the restricted predictiors is discussed. Simulations are used to demonstrate that imposing a restriction can reduce the bias compare to that of the small area predictors without restriction.

## 1. Introduction

There are many situations in which it is desirable to derive reliable estimators for small geographical areas or small subpopulations from existing survey data. However, sample sizes for these small areas are typically small due to their relative size. Therefore, the usual direct survey estimators for such small areas, based on data only from the sample units in the area, are likely to yield unacceptably large standard errors (compared to the interesting statistic). This makes it necessary to "borrow strength" from related area using a model-dependent estimator to find more accurate estimates for the given area or, simultaneously, for several areas.

The random effects model for small area estimation is

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + z_i\, b_i, \tag{1}$$

and

$$Y_i = y_i + e_i \qquad i = 1, \cdots, n, \tag{2}$$

where $y_i$ are unobservable small area means of interest, $Y_i$ are observable direct survey estimators, $\boldsymbol{x}_i^T$ are known constants (auxillary information), $z_i$ are known positive constants, $\boldsymbol{\beta}$ is the vector of regression parameters, the $b_i$'s are independent and identically distributed random variables with $E(b_i) = 0$ and $V(b_i) = \sigma_b^2$, and the $e_i$'s are sampling errors with $E(e_i|y_i) = 0$ and $V(e_i|y_i) = \sigma_{ei}^2$. Small area estimation is the construction of predictors for $y_i$ and the estimation of the $MSE$ of the small area predictors.

Combining (1) and (2), we obtain the model

$$Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + z_i\, b_i + e_i, \qquad i = 1, \ldots, n, \tag{3}$$

which is a special case of the general mixed linear model. When the variance components are known, the best linear unbiased predictor (BLUP) of $y_i$ is

$$\widetilde{y}_i^H = \gamma_i Y_i + (1 - \gamma_i) \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}, \tag{4}$$

where $\widehat{\boldsymbol{\beta}} = \left[\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right]^{-1} \boldsymbol{X}'\boldsymbol{V}^{-1}\mathbf{Y}$, and

$$\gamma_i = z_i^2 \sigma_b^2 (z_i^2 \sigma_b^2 + \sigma_{ei}^2)^{-1}. \tag{5}$$

See Henderson (1963). This result does not depend on normality for $b_i$ and $e_i$.

We use $\widetilde{\boldsymbol{y}}^H$ instead of the direct survey estimator $Y_i$ to estimate $y_i$ for each small area $i$ because $Y_i$ has a large variance as an estimator of $y_i$ due to small sample size. However, the direct survey estimator of the mean across all (or several) small areas is often satisfactory when the sample size is large. For example, the direct survey estimator in Fuller and Wang (2000) for urban change acres $Y_i$ has a large variance as an estimator of the true urban change acres $y_i$ in the $i$-th small area (HUCCO). However, the direct survey estimator for state urban change acres $\sum_i Y_i$ is design unbiased for the true urban state change acres, and it has relatively small variance. Therefore, it is often desirable to put a restriction on the weighted total of the small area estimators such that the weighted total of the small area estimators is equal to the the weighted total of the direct survey estimators.

This research was funded in part by cooperative agreement 68-3A75-14 between the USDA Natural Resources Conservation Services and Iowa State University

Equivalently, we can put the restriction on the weighted mean of the small area estimators. Thus, we want to modify the small area estimators such that

$$\sum_{i=1}^{n} \omega_i \widehat{y}_i^M = \sum_{i=1}^{n} \omega_i Y_i, \tag{6}$$

where $\omega_i > 0, i = 1, \cdots, n$, are the weights, $\sum_{i=1}^{n} \omega_i = 1$, and $\widehat{y}_i^M$ is the adjusted small area estimator. Usually the $\omega_i$ are the sampling weights such that $\sum_{i=1}^{n} \omega_i Y_i$ is an unbiased estimator of the population mean. One heuristic approach is to make a ratio adjustment to obtain

$$\widehat{y}_i^M = \left( \sum_{j=1}^{n} \omega_j \widehat{y}_j^H \right)^{-1} \left( \sum_{j=1}^{n} \omega_j Y_j \right) \widehat{y}_i^H, \tag{7}$$

where $\widehat{y}_i^H$ is the EBLUP of $y_i$. This adjustment was used in Fuller and Wang (2000). The disadvantage of this approach is that it is difficult to assess the bias and variance of $\widehat{y}_i^M$.

Pfeffermann and Barnard (1991) proposed an alternative approach. The mixed model equation for the random effects model defined in (3) is

$$\left[ \begin{array}{cc} \boldsymbol{X}'\boldsymbol{\Sigma}_e^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{\Sigma}_e^{-1}\mathbf{Z} \\ \mathbf{Z}'\boldsymbol{\Sigma}_e^{-1}\boldsymbol{X} & \mathbf{Z}'\boldsymbol{\Sigma}_e^{-1}\mathbf{Z} + \boldsymbol{\Sigma}_b^{-1} \end{array} \right] \left[ \begin{array}{c} \boldsymbol{\beta} \\ \mathbf{b} \end{array} \right] = \left[ \begin{array}{c} \boldsymbol{X}'\boldsymbol{\Sigma}_e^{-1}\mathbf{Y} \\ \mathbf{Z}'\boldsymbol{\Sigma}_e^{-1}\mathbf{Y} \end{array} \right], \tag{8}$$

where $\boldsymbol{\Sigma}_b = diag(z_1^2\sigma_b^2, \cdots, z_n^2\sigma_b^2)$ and $\boldsymbol{\Sigma}_e = diag(\sigma_{e1}^2, \cdots, \sigma_{en}^2)$. Let $\widetilde{\boldsymbol{y}}^H = (\widetilde{y}_1^H, \cdots, \widetilde{y}_n^H)^T$ denote the BLUP estimator of $\boldsymbol{y} = (y_1, \cdots, y_n)^T$. We have

$$\widetilde{\boldsymbol{y}}^H = \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \mathbf{Z}\widehat{\boldsymbol{b}}, \tag{9}$$

where $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b}}$ are any solutions to the mixed model equation (8). Note that finding a solution to the mixed model equation (8) is equivalent to finding a solution to the minimization problem

$$\min_{\boldsymbol{\beta}, \boldsymbol{b}} \ \{(\boldsymbol{Y} - \boldsymbol{X\beta} - \boldsymbol{Zb})^T \boldsymbol{\Sigma}_e^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta} - \boldsymbol{Zb}) + \boldsymbol{b}^T\boldsymbol{\Sigma}_b^{-1}\boldsymbol{b}\Big\}. \tag{10}$$

To make (6) hold, Pfeffermann and Barnard (1991) proposed the modified estimator

$$\widehat{\boldsymbol{y}}^M = \boldsymbol{X}\widehat{\boldsymbol{\beta}}^M + \mathbf{Z}\widehat{\boldsymbol{b}}^M, \tag{11}$$

where $\widehat{\boldsymbol{\beta}}^M$ and $\widehat{\boldsymbol{b}}^M$ are any solutions to the minimization problem (10) with $\boldsymbol{\beta}$ and $\boldsymbol{b}$ subject to the constraint

$$\sum_{i=1}^{n} \omega_i(\boldsymbol{x}_i^T\boldsymbol{\beta} + z_i b_i) = \sum_{i=1}^{n} \omega_i Y_i. \tag{12}$$

This leads to the estimator

$$\widehat{y}_i^M = \widetilde{y}_i^H + [Var(\widetilde{y}_.)]^{-1}cov(\widetilde{y}_i^H, \widetilde{y}_.)[\sum_{j=1}^{n} \omega_j Y_j - \widetilde{y}_.], \tag{13}$$

where $\widetilde{y}_. = \sum_{i=1}^{n} \omega_i \widetilde{y}_i^H$. Pfeffermann and Barnard did not give the expression for $cov(\widetilde{y}_i^H, \widetilde{y}_.)$ and $Var(\widetilde{y}_.)$ in their paper.

The Pfeffermann-Barnard (1991) approach is a natural way to make the estimator $\widehat{y}_i^M$ satisfy (6). The derivation of (13) relies on the fact that $\boldsymbol{b}$ can be treated as a fixed parameter. When there is no restriction, Henderson (1950) showed that we can solve the mixed model equation (8) and estimate $\boldsymbol{b}$ as a fixed parameter. However, (12) puts a constraint on the random vector $\boldsymbol{b}$. The constraint (12) makes the distribution of $\boldsymbol{b}$ a degenerate one. Thus, the variance structure of $\boldsymbol{b}$ is changed and $Var(\boldsymbol{b}) \neq diag(z_1^2\sigma_b^2, \ldots, z_n^2\sigma_b^2)$. The underlying model assumptions about the random effects model (1) have been changed by the constraint. It is not clear that the estimation of $\boldsymbol{b}$ as a fixed parameter is still justified.

Isaki, Tsay and Fuller (2000) imposed the restriction by a procedure that, approximately, constructed the best predictors of $n-1$ quantities that are estimated to be uncorrelated with $\sum_{i=1}^{n} \omega_i Y_i$. Let

$$\boldsymbol{\omega} = (\omega_1, \cdots, \omega_n)^T, \tag{14}$$

and

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_b + \boldsymbol{\Sigma}_e = Var(\boldsymbol{Y}). \tag{15}$$

Let $\widehat{\boldsymbol{\Sigma}}$ be the estimator of $\boldsymbol{\Sigma}$ and let $\ddot{\boldsymbol{C}} = \ddot{\boldsymbol{A}}\boldsymbol{T}$, where

$$\boldsymbol{T} = \left( \begin{array}{cc} \multicolumn{2}{c}{\boldsymbol{\omega}'} \\ \mathbf{0}_{n-1} & \boldsymbol{I}_{n-1} \end{array} \right), \tag{16}$$

$$\ddot{\boldsymbol{A}} = \left( \begin{array}{cc} 1 & \mathbf{0}'_{n-1} \\ -\ddot{\boldsymbol{a}}_{n-1} & \boldsymbol{I}_{n-1} \end{array} \right),$$

$$\ddot{\boldsymbol{a}}_{n-1} = \left(\boldsymbol{\omega}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{\omega}\right)^{-1} \left( \ \mathbf{0}_{n-1} \quad \boldsymbol{I}_{n-1} \ \right) \widehat{\boldsymbol{\Sigma}}\boldsymbol{\omega}.$$

The modified estimator of $\boldsymbol{y}$ is

$$\widehat{\boldsymbol{y}}^M = \boldsymbol{Y} - \ddot{\boldsymbol{C}}^{-1}\boldsymbol{B}\ddot{\boldsymbol{C}}(\boldsymbol{I}_n - \widehat{\boldsymbol{\Gamma}})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}), \tag{17}$$

where $\widehat{\boldsymbol{\Gamma}} = \widehat{\boldsymbol{\Sigma}}_b\widehat{\boldsymbol{\Sigma}}^{-1}$, and

$$\boldsymbol{B} = \left( \begin{array}{cc} 0 & \mathbf{0}'_{n-1} \\ \mathbf{0}_{n-1} & \boldsymbol{I}_{n-1} \end{array} \right). \tag{18}$$

Isaki et. al. argued that the estimator in (17) gave the BLUP of the $n-1$ quantities orthogonal to $\sum_{i=1}^{n} \omega_i Y_i$

when variance components parameters are known, but no other theoretical justification was provided for the particular choice of $\ddot{\boldsymbol{a}}_{n-1}$.

In light of this fact, it is desirable to develope theory for small areas estimation under restriction (6). We consider restriction (6) as an adjustment problem instead of a constraint problem proposed by Pfeffermann and Barnard. Suppose we have the small area estimator $\widehat{y}_i$. To make the modified (adjusted) estimator $\widehat{y}_i^M$ satisfy (6), we construct $\widehat{y}_i^M$ by allocating the difference $\sum_{j=1}^n \omega_j Y_j - \sum_{j=1}^n \omega_j \widehat{y}_j$ to the small area estimators $\widehat{y}_i$, $i = 1, \cdots, n$, according to a rule defined by

$$\widehat{y}_i^M = \widehat{y}_i + a_i \left[ \sum_{j=1}^n \omega_j Y_j - \sum_{j=1}^n \omega_j \widehat{y}_j \right], \tag{19}$$

where $\sum_{i=1}^n \omega_i a_i = 1$. Clearly, the modified estimator satisfies (6).

The estimator defined in (7) is an estimator of the form (19) since

$$\widehat{y}_i^M = \widehat{y}_i^H + a_i \left( \sum_{j=1}^n \omega_j Y_j - \sum_{j=1}^n \omega_j \widehat{y}_j^H \right), \tag{20}$$

where $a_i = \left( \sum_{j=1}^n \omega_j \widehat{y}_j^H \right)^{-1} \widehat{y}_i^H$. The estimator defined in (13) is also of the form (19) though the estimator is derived from a constraint minimization problem. An estimator similar to (13) proposed by Battese, Harter and Fuller (1988) is

$$\widehat{y}_i^M = \widehat{y}_i^H + a_i \left( \sum_{j=1}^n \omega_j Y_j - \sum_{j=1}^n \omega_j \widehat{y}_j^H \right), \tag{21}$$

where $a_i = \left[ \sum_{j=1}^n \omega_j^2 \widehat{Var}(\widehat{y}_j^H) \right]^{-1} \omega_i \widehat{Var}(\widehat{y}_i^H)$.

For the random effects model defined in (1) and (2), $\widehat{\boldsymbol{\Sigma}} = diag(z_1^2 \widehat{\sigma}_b^2 + \widehat{\sigma}_{e1}^2, \cdots, z_n^2 \widehat{\sigma}_b^2 + \widehat{\sigma}_{en}^2)$, $\widehat{\boldsymbol{\Gamma}} = diag(\widehat{\gamma}_1, \cdots, \widehat{\gamma}_n)$, and $\widehat{\gamma}_i = (z_i^2 \widehat{\sigma}_b^2 + \widehat{\sigma}_{ei}^2)^{-1} z_i^2 \widehat{\sigma}_b^2$. After some matrix operation, we can rewrite (17) in component form, rather than matrix form, as

$$\widehat{y}_i^M = \widehat{y}_i^H + a_i \left( \sum_{j=1}^n \omega_j Y_j - \sum_{j=1}^n \omega_j \widehat{y}_j^H \right), \tag{22}$$

where $a_i = \left[ \sum_{j=1}^n \omega_j^2 \widehat{Var}(Y_j) \right]^{-1} \omega_i \widehat{Var}(Y_i)$. Therefore, the Isaki, Tsay and Fuller estimator in (17) also has the form of (19). We will discuss the properties of the modified estimator of the form (19) in this paper.

## 2. Best linear unbiased estimator under a restriction

We want to find the "best" linear unbiased estimator for $\boldsymbol{y}$ that satisfies restriction (6). Just as in the derivation of BLUP, we first assume the parameters for the variance components are known. Let $R(\widehat{\boldsymbol{y}})$ denote the collection of all linear unbiased estimators that satisfies (6). Suppose the BLUP of $\boldsymbol{y} = (y_1, \cdots, y_n)^T$ is $\widetilde{\boldsymbol{y}}^H = (\widetilde{y}_1^H, \cdots, \widetilde{y}_n^H)^T$ and $\widetilde{\boldsymbol{y}}^H \notin R(\widehat{\boldsymbol{y}})$.

First, we need to define the meaning of "best". Given the constraint (6), we cannot obtain the BLUP for all $y_i, i = 1, \ldots, n$. Consider a family of estimators $\widetilde{\boldsymbol{y}}^{(k)}$, where

$$\widetilde{y}_i^{(k)} = \begin{cases} \widetilde{y}_i^H & \text{if } i \neq j, \\ \widetilde{y}_i^H + \omega_i^{-1} \left[ \sum_{j=1}^n \omega_j Y_j - \sum_{j=1}^n \omega_j \widetilde{y}_j^H \right] & \text{if } i = k. \end{cases} \tag{23}$$

In other words, $\widetilde{\boldsymbol{y}}^{(k)}$ is the estimator in which every component is the BLUP of $y_i$ except the $k$-th component is the BLUP of $y_k$ plus $\omega_k^{-1} \left[ \sum_{j=1}^n \omega_j Y_j - \sum_{j=1}^n \omega_j \widetilde{y}_j^H \right]$. It is easy to see that $\widetilde{\boldsymbol{y}}^{(k)} \in R(\widehat{\boldsymbol{y}})$. For any $\widehat{\boldsymbol{y}}^M \in R(\widehat{\boldsymbol{y}})$, there is at least one component $\widehat{y}_i^M \neq \widetilde{y}_i^H$ since $\widetilde{\boldsymbol{y}}^H \notin R(\widehat{\boldsymbol{y}})$. For any $\widetilde{\boldsymbol{y}}^{(k)}$, where $k \neq i$,

$$Var(\widehat{y}_i^M - y_i) > Var(\widetilde{y}_i^{(k)} - y_i) = Var(\widetilde{y}_i^H - y_i) \tag{24}$$

because $\widetilde{y}_i^H$ is the BLUP of $y_k$. Therefore, for any $\widehat{\boldsymbol{y}}^M \in R(\widehat{\boldsymbol{y}})$, there are always at least another $n-1$ estimators in $R(\widehat{\boldsymbol{y}})$ with smaller prediction variance for at least one component. This indicates that no estimator can be found in $R(\widehat{\boldsymbol{y}})$ with smallest prediction variance for every component.

Since it is impossible to compare estimators in $R(\widehat{\boldsymbol{y}})$ component-by-component to find the best estimator, some kind of overall criterion is desirable. A natural choice is to find $\widehat{\boldsymbol{y}}^M \in R(\widehat{\boldsymbol{y}})$ that minimizes

$$Q(\widehat{\boldsymbol{y}}^M) = \sum_{i=1}^n \varphi_i E(\widehat{y}_i^M - y_i)^2, \tag{25}$$

where the $\varphi_i, i = 1, \ldots, n$ are positive weights. Usually, $\varphi_i$ depends on the variance components. To obtain a general result, we do not specify $\varphi_i$ now. We will discuss the choice of $\varphi_i$ later. We give the major theorem of this paper.

**Theorem 1** *Assume the random effects model*

$$Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + z_i b_i + e_i, \qquad i = 1, \ldots, n,$$

*where the $b_i$ have independent identical distributions with mean zero and variance $\sigma_b^2$, the $e_i$ have independent distributions with mean zero and variance $\sigma_{ei}^2$, and $\boldsymbol{b} = (b_1, \ldots, b_n)^T$ is independent of $\boldsymbol{e} = (e_1, \ldots, e_n)^T$. Assume $z_i$, $\sigma_b^2$, and $\sigma_{ei}^2$ are known and $\boldsymbol{\beta}$ is unknown. Let $\widetilde{y}_i^H$ be the BLUP of $y_i$ defined in (4). Let*

$$\widehat{y}_i^M = \widetilde{y}_i^H + \breve{a}_i \left( \sum_{j=1}^n \omega_j Y_j - \sum_{j=1}^n \omega_j \widetilde{y}_j^H \right), \tag{26}$$

*where $\breve{a}_i = \left( \sum_{i=1}^n \varphi_i^{-1} \omega_i^2 \right)^{-1} \varphi_i^{-1} \omega_i$ and $\omega_i$ are the fixed weights of (6). Then $\widehat{\boldsymbol{y}}^M = (\widetilde{y}_1^M, \ldots, \widetilde{y}_n^M)^T$ is the* **unique** *estimator among all linear unbiased estimators that satisfies (6) and minimizes criterion (25). In other words, $\widehat{\boldsymbol{y}}^M$ is the unique best estimator in $R(\widehat{\boldsymbol{y}})$ under criterion (25).*

**Proof:** See Appendix A.

Remark 1. When the variance components are unknown, we replace the variance components in (5) with estimators to obtain the empirical BLUP or EBLUP, denoted by $\widehat{y}_i^H$. To impose the restriction (6), we define the modified estimator

$$\widehat{y}_i^M = \widehat{y}_i^H + \hat{a}_i (\sum_{j=1}^n \omega_j Y_j - \sum_{j=1}^n \omega_j \widehat{y}_j^H), \tag{27}$$

where $\widehat{y}_i^H$ is the EBLUP.

Remark 2. If $\varphi_i = \omega_i Y_i^{-1}$, we have that the ratio estimator defined in (20) minimizes the criterion (25). Since $Y_i$ could be less than zero and it is not reasonable to have negative $\varphi_i$, we can see that the ratio adjustment is not always a good choice. If $\varphi_i = \omega_i [cov(\widetilde{y}_i^H, \widetilde{y}_.)]^{-1}$, where $\widetilde{y}_. = \sum_{j=1}^n \omega_j \widetilde{y}_j^H$, we have the estimator (13) derived by Pfeffermann and Barnard (1991). The estimators in (21) and (22) are estimators when the variance components are unknown. When $\varphi_i = [\widehat{Var}(\widehat{y}_j^H)]^{-1}$, we have the Battese, Harter and Fuller estimator of (21). The Isaki, Tsay and Fuller estimator in (22) results from $\varphi_i = [\widehat{Var}(Y_i)]^{-1}$. Therefore, Theorem 1 provides a unified way to derive the estimators described in the introduction part of this paper.

## 3. Choice of Criteria

Since the "best" estimator depends on the $\varphi_i$ used in the criterion (25), we desire a reasonable choice of $\varphi_i$. To gain insight into the problem, we first assume all the variance components to be known.

We argue that $\varphi_i = [Var(Y_i)]^{-1}$ is the most reasonable choice by showing the properties of the corresponding modified estimator. Let

$$\dot{\boldsymbol{a}}_{n-1} = (\dot{a}_2, \ldots, \dot{a}_n)' = (\boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega})^{-1} \begin{pmatrix} \boldsymbol{0}_{n-1} & \boldsymbol{I}_{n-1} \end{pmatrix} \boldsymbol{\Sigma} \boldsymbol{\omega},$$

where $\boldsymbol{\omega}$ is defined in (14) and $\boldsymbol{\Sigma}$ is defined in (15). Let $\dot{\boldsymbol{C}} = \dot{\boldsymbol{A}} \boldsymbol{T}$, where

$$\dot{\boldsymbol{A}} = \begin{pmatrix} 1 & 0'_{n-1} \\ -\dot{\boldsymbol{a}}_{n-1} & \boldsymbol{I}_{n-1} \end{pmatrix}$$

and $\boldsymbol{T}$ is defined in (16). We want to estimate $\boldsymbol{y}$. Equivalently, we can estimate $\dot{\boldsymbol{C}}\boldsymbol{y} = (\bar{y}, y_2 - \dot{a}_2 \bar{y}, \ldots, y_n - \dot{a}_n \bar{y})'$, where $\bar{y} = \sum_{i=1}^n \omega_i y_i$. When there is no restriction, the BLUP of $\dot{\boldsymbol{C}}\boldsymbol{y}$ is $\dot{\boldsymbol{C}}\widetilde{\boldsymbol{y}}^H$ by the property of BLUP. In other words,

$$\widetilde{\dot{\boldsymbol{C}}\boldsymbol{y}} = \dot{\boldsymbol{C}}\boldsymbol{Y} - \dot{\boldsymbol{C}}(\boldsymbol{I}_n - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}). \tag{28}$$

Let $\bar{Y} = \sum_{i=1}^n \omega_i Y_i$ and observe that $\dot{\boldsymbol{C}}\boldsymbol{Y} = (\bar{Y}, Y_2 - \dot{a}_2 \bar{Y}, \ldots, Y_n - \dot{a}_n \bar{Y})'$. Note that $Y_i - \dot{a}_i \bar{Y}, i = 2, \ldots, n$ are uncorrelated with $\bar{Y}$, i.e., $(Y_2 - \dot{a}_2 \bar{Y}, \ldots, Y_n - \dot{a}_n \bar{Y})'$ is a basis for the space that is orthogonal to $\bar{Y}$ in the space spanned by $\boldsymbol{Y}$. Let $\dot{\boldsymbol{c}}_1'$ be the first row of $\dot{\boldsymbol{C}}$, $\dot{\boldsymbol{c}}_1' \boldsymbol{y} = \bar{y} = \sum_{i=1}^n \omega_i y_i$. To impose the restriction (6) on $\dot{\boldsymbol{C}}\boldsymbol{y}$, we only need to replace $\widetilde{\dot{\boldsymbol{c}}_1' \boldsymbol{y}}$ with $\sum_{i=1}^n \omega_i Y_i$ and use BLUP to estimate the other $n-1$ quantities in $\dot{\boldsymbol{C}}\boldsymbol{y}$. This leads to the modified estimator

$$\dot{\boldsymbol{C}}\widetilde{\boldsymbol{y}}^M = \dot{\boldsymbol{C}}\boldsymbol{Y} - \boldsymbol{B}\dot{\boldsymbol{C}}(\boldsymbol{I}_n - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}), \tag{29}$$

where $\boldsymbol{B}$ is defined in (18). To obtain the modified estimator for $\boldsymbol{y}$, we multiply equation (29) by $\dot{\boldsymbol{C}}^{-1}$ on both sides to obtain

$$\widetilde{\boldsymbol{y}}^M = \boldsymbol{Y} - \dot{\boldsymbol{C}}^{-1} \boldsymbol{B} \dot{\boldsymbol{C}} (\boldsymbol{I}_n - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}). \tag{30}$$

This estimator does not depend on the choice of the basis of the subspace (of the space spanned by $\boldsymbol{Y}$) that is orthogonal to $\bar{Y}$. Therefore, (30) is derived by applying the restriction and making estimation based on the information from the space that is orthogonal to $\bar{Y}$.

There are exactly $n-1$ linearly independent BLUPs in any $n-1$ dimensional subspace spanned by $\boldsymbol{Y}$ that does not contain $\bar{Y} = \sum_{i=1}^n \omega_i Y_i$. Let $\boldsymbol{C}_{n-1}\boldsymbol{Y}$ be the space spanned by the $n-1$ linearly independent BLUPs and let $\boldsymbol{C_a} = \left(\boldsymbol{\omega}, \boldsymbol{C}'_{n-1}\right)'$. Similar to the derivation of $\tilde{\boldsymbol{y}}^M$ in (30), we can impose the restriction (6) by using the linear combination of $\sum_{i=1}^{n} \omega_i Y_i$ and the $n-1$ BLUPs to construct the modified estimator

$$\widetilde{\boldsymbol{y}_{\boldsymbol{a}}}^M = \boldsymbol{Y} - \boldsymbol{C}_{\boldsymbol{a}}^{-1} \boldsymbol{B} \boldsymbol{C}_{\boldsymbol{a}} (\boldsymbol{I}_n - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}). \tag{31}$$

It is possible to derive some linear unbiased estimators for $\boldsymbol{y}$ by using less than $n-1$ linearly independent BLUPs in the space spanned $\boldsymbol{Y}$. Obviously, these estimators are less efficient than the estimators in the form of (31). Therefore, the question, "which estimator for $\boldsymbol{y}$ is the most reasonable estimator?" is equivalent to the question, "which $n-1$ BLUPs are the most reasonable choice for constructing the restricted estimator?" If we interpret restriction (6) as meaning that $\bar{Y} = \sum_{i=1}^{n} \omega_i Y_i$ is the best estimator for $\bar{y} = \sum_{i=1}^{n} \omega_i y_i$, we should choose to evaluate the $n-1$ components that are orthogonal to $\bar{Y}$. This indicates that the estimator in (30) is the most reasonable estimator. Therefore, $\varphi_i = [Var(Y_i)]^{-1}$ is the most sensible choice for $\varphi_i$. When the variance components are unknown, we replace the variance components with the corresponding estimated value. This leads to the estimator in (22).

We further establish the relationship between the estimators of the form (31) and estimator of the form (19). Let $\boldsymbol{a}$ be the first column of $\boldsymbol{C}_{\boldsymbol{a}}^{-1}$ and let

$$\widehat{\boldsymbol{y}} = (\widehat{y}_1, \ldots, \widehat{y}_n)' = [\boldsymbol{Y} - (\boldsymbol{I}_n - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})]$$

be a small area estimator of $\boldsymbol{y}$. Then

$$\begin{aligned} \tilde{\boldsymbol{y}}_{\boldsymbol{a}}^M &= \boldsymbol{Y} - \boldsymbol{C}_{\boldsymbol{a}}^{-1} \boldsymbol{B} \boldsymbol{C}_{\boldsymbol{a}} (\boldsymbol{I}_n - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\ &= [\boldsymbol{Y} - (\boldsymbol{I}_n - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})] \qquad (32) \\ &\quad -\boldsymbol{C}_{\boldsymbol{a}}^{-1} (\boldsymbol{I}_n - \boldsymbol{B}) C_{\boldsymbol{a}} (\boldsymbol{I}_n - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\ &= \widehat{\boldsymbol{y}} - \boldsymbol{a}[\sum_{j=1}^{n} \omega_j Y_j - \sum_{j=1}^{n} \omega_j \widehat{y}_j] \qquad (33) \end{aligned}$$

because $\boldsymbol{C}_{\boldsymbol{a}}^{-1}(\boldsymbol{I}_n - \boldsymbol{B})C_{\boldsymbol{a}} = [\boldsymbol{C}_{\boldsymbol{a}}^{-1}(\boldsymbol{I}_n - \boldsymbol{B})][(\boldsymbol{I}_n - \boldsymbol{B})C_{\boldsymbol{a}}] = \boldsymbol{a}\boldsymbol{\omega}'$. In other words, any estimator of the form (31) can be written in the form of (19). On the other hand, any estimator of the form (19) can be written in the form of (31) by letting $\boldsymbol{C_a} = \boldsymbol{A_a T}$, where

$$\boldsymbol{A_a} = \begin{pmatrix} 1 & \boldsymbol{0}'_{n-1} \\ -\boldsymbol{a}_{n-1} & \boldsymbol{I}_{n-1} \end{pmatrix},$$

$$\boldsymbol{a}_{n-1} = (a_2, \ldots, a_n)'.$$

Therefore, (19) and (31) are different representations of the same family of estimators.

## 4. References


Battese, George E., Harter, Rachel M. and Fuller, Wayne A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of American Statistical Association*, **83**, 28-36.

Henderson, C. R. (1950). Estimation of genetic parameters (abstract). *Annals of Mathematical Statistics*, **21**, 309-310.

Henderson, C. R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding*, (eds Hanson, W.D. and Robinson, H.F.), 141-163. National Academy of Sciences–National Research Council, Washington. Publication 982.

Isaki, C. T., Tsay, J. H., and Fuller, W. A. (2000). Estimation of Census Adjustment Factors, *Survey Methodology* **26**, 31-42.

Fuller, W. A. and Wang, J. (2000). Geographic information in small area estimation for a National survey, *Statistics in Transition* **4**, 587-596.

Pfeffermann, D. and Barnard, C. H. (1991). Some new estimators for small-area means with application to the assessment of Farmland values, *Journal of Business & Economic Statistics* **9**, 73-84.


## 5. Appendix

**Proof of Theorem 1** Let $R^0(\tilde{\boldsymbol{y}}_{\boldsymbol{a}})$ denote the collection of all estimators that have the form

$$\tilde{y}_{i\boldsymbol{a}}^M = \tilde{y}_i^H + a_i \left( \sum_{j=1}^{n} \omega_j Y_j - \sum_{j=1}^{n} \omega_j \tilde{y}_j^H \right), \tag{34}$$

where $\sum_{i=1}^{n} \omega_i a_i = 1$. Clearly, $R^0(\tilde{\boldsymbol{y}}_{\boldsymbol{a}})$ is a subset of $R(\widehat{\boldsymbol{y}})$. We first find the best estimator in $R^0(\tilde{\boldsymbol{y}}_{\boldsymbol{a}})$. For any $\tilde{\boldsymbol{y}}_{\boldsymbol{a}}^M \in R^0(\tilde{\boldsymbol{y}}_{\boldsymbol{a}})$, let

$$\begin{aligned} f(a_1, \ldots, a_n) = Q(\tilde{\boldsymbol{y}}_{\boldsymbol{a}}^M) &= \sum_{i=1}^{n} \varphi_i E\{[(\tilde{y}_i^H - y_i) \\ &\quad + a_i(\sum_{j=1}^{n} \omega_j Y_j - \sum_{j=1}^{n} \omega_j \tilde{y}_j^H)]^2\}. \end{aligned}$$

Note that $cov\left\{(1-\gamma_i)\boldsymbol{x}_i^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \omega_j(1-\gamma_j)\right\} = 0$ and $cov\{(\gamma_i - 1)b_i + \gamma_i e_i,\, \omega_j(1-\gamma_j)[(b_j + e_j) - \boldsymbol{x}_j^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})]\} = 0$, where $\gamma_i$ is defined in (5). This leads to

$$E\left[(\tilde{y}_i^H - y_i)(Y_j - \tilde{y}_j^H)\right] = 0 \tag{35}$$

for $i = 1, ..., n$ and $j = 1, ..., n$. Therefore,

$$f(a_1, \ldots, a_n) = \sum_{i=1}^{n} \varphi_i E\left\{(\tilde{y}_i^H - y_i)^2\right\}$$

$$+E\left\{\left[\sum_{j=1}^{n}\omega_j(Y_j-\widetilde{y}_j^H)\right]^2\right\}\sum_{i=1}^{n}\varphi_i a_i^2. \quad (36)$$

Using Lagrangian multiplier methods to minimize $f(a_1,\ldots,a_n)$ subject to the restriction $\sum_{i=1}^{n}\omega_i a_i=1$, we obtain the system

$$2a_i\varphi_i E\left\{\left[\sum_{j=1}^{n}\omega_j(Y_j-\widetilde{y}_j^H)\right]^2\right\}+\lambda\omega_i=0, i=1,\ldots,n.$$

The solution to the linear system subject to the restriction $\sum_{i=1}^{n}\omega_i a_i=1$ is

$$\breve{a}_i=\left(\sum_{i=1}^{n}\varphi_i^{-1}\omega_i^2\right)^{-1}\varphi_i^{-1}\omega_i. \quad (37)$$

Therefore, $\widehat{\boldsymbol{y}}^M$ defined in (26) is an estimator of the form (34) and minimizes criterion (25).

Let $\widehat{\boldsymbol{y}}$ be any linear unbiased estimator of $\boldsymbol{y}$ that satisfies (6), i.e., $\widehat{\boldsymbol{y}}\in R(\widehat{\boldsymbol{y}})$. By standard results for BLUP (See, for example, Robinson (1991) and Harville (1976)), we have

$$cov(\widetilde{y}_i^H-y_i,\widehat{y}_i-\widetilde{y}_i^H)=0.$$

This leads to

$$E\left\{(\widehat{y}_i-y_i)^2\right\}=E\left\{(\widetilde{y}_i^H-y_i)^2\right\}+E\left\{(\widehat{y}_i-\widetilde{y}_i^H)^2\right\}. \quad (38)$$

Therefore,

$$Q(\widehat{\boldsymbol{y}})=\sum_{i=1}^{n}\varphi_i E\left\{(\widetilde{y}_i^H-y_i)^2\right\}+\sum_{i=1}^{n}\varphi_i E\left\{(\widehat{y}_i^M-\widetilde{y}_i^H)^2\right\}. \quad (39)$$

Since $\widehat{\boldsymbol{y}}$ satisfies (6), we have $\sum_{i=1}^{n}\omega_i\widehat{y}_i=\sum_{i=1}^{n}\omega_i Y_i$. For the $\widehat{\boldsymbol{y}}^M$ defined in (26),

$$\widehat{y}_i^M=\widetilde{y}_i^H+\breve{a}_i\left[\sum_{j=1}^{n}\omega_j(\widehat{y}_j-\widetilde{y}_j^H)\right].$$

By (36), we have

$$Q(\widehat{\boldsymbol{y}}^M)=\sum_{i=1}^{n}\varphi_i E\left\{(\widetilde{y}_i^H-y_i)^2\right\}$$
$$+E\left\{\left[\sum_{j=1}^{n}\omega_j(\widehat{y}_j-\widetilde{y}_j^H)\right]^2\right\}\left(\sum_{i=1}^{n}\varphi_i^{-1}\omega_i^2\right)^{-1} \quad (40)$$

Note that

$$E\left\{\left[\sum_{j=1}^{n}\omega_j(\widehat{y}_j-\widetilde{y}_j^H)\right]^2\right\}\leq\sum_{j=1}^{n}\sum_{k=1}^{n}\omega_j\omega_k g_j g_k = \left(\sum_{i=1}^{n}\omega_i g_i\right)^2, \quad (41)$$

where $g_j=\sqrt{E\left\{(\widehat{y}_j-\widetilde{y}_j^H)^2\right\}}$. By Cauchy's inequality,

$$\left(\sum_{j=1}^{n}\omega_j g_j\right)^2\leq\left(\sum_{i=1}^{n}\varphi_i^{-1}\omega_i^2\right)\left(\sum_{i=1}^{n}\varphi_i g_i^2\right) = \left(\sum_{i=1}^{n}\varphi_i^{-1}\omega_i^2\right)\left[\sum_{i=1}^{n}\varphi_i E\left\{(\widehat{y}_j-\widetilde{y}_j^H)^2\right\}\right]. \quad (42)$$

Combining (40), (41) and (42), we have $Q(\widehat{\boldsymbol{y}}^M)\leq Q(\widehat{\boldsymbol{y}})$.

To show the uniqueness of $\widehat{\boldsymbol{y}}^M$, we need to check when the inequalities (41) and (42) become equalities. Inequality (41) becomes an equality if and only if

$$\widehat{y}_j-\widetilde{y}_j^H=c_j^0+c_j^1(\widehat{y}_1-\widetilde{y}_1^H) \quad (43)$$

for some constants $c_j^0$ and $c_j^1, j=2,\ldots,n$. Inequality (42) becomes an equality if and only if

$$\sqrt{\varphi_i^{-1}\omega_i^2}\sqrt{v_j g_j^2}-\sqrt{v_j^{-1}\omega_j^2}\sqrt{\varphi_i g_i^2}=0,$$

or, equivalently,

$$\varphi_i^2\omega_i^{-2}E\left\{(\widehat{y}_i-\widetilde{y}_i^H)^2\right\}=v_j^2\omega_j^{-2}E\left\{(\widehat{y}_j-\widetilde{y}_j^H)^2\right\}. \quad (44)$$

Also,

$$\sum_{i=1}^{n}\omega_i\widehat{y}_i=\sum_{i=1}^{n}\omega_i Y_i. \quad (45)$$

Combining (43), (44), and (45), we have that the equality holds if and only if $\widehat{y}_j=\widehat{y}_j^M$. Thus, we have shown that $\widehat{\boldsymbol{y}}^M$ is the unique linear unbiased estimator that satisfies (6) and minimizes criterion (25).