Journal of Official Statistics
Vol. 7. No. 1, 1991, pp. 69–76

A Model-Based Approach: Composite Estimators for Small Area Estimation

Kung-Jong Lui[1] *and William G. Cumberland*[2]

Abstract: To reduce the mean-squared-error (MSE) of an estimator in small area estimation, the composite estimator, a weighted sum of two component estimators, is often considered. The difficulties of providing a measure of error with respect to the sampling plan for this estimator, and of deriving the optimal weight to minimize MSE, limit its use. We propose a super-population model-based approach to derive explicit optimal weights for the composite estimator under several models related to the synthetic estimator. In addition, the prediction variance of the composite estimator is easily obtained. A simple test to help in deciding how to best apply these results to small area estimation is given.

Key words: Composite estimator; optimal weight; super-population.

1. Introduction

In small area estimation, samples designed to provide estimates for large geographic areas are often used to provide estimates for small areas as well. In such cases the sample in a small area may be unrepresentative or too small to produce reliable estimates; synthetic estimators are often suggested for these situations. The composite estimator, a weighted sum of two component estimators, can have a mean-squared-error (MSE) which is smaller than that of either component estimator when an appropriate weighting scheme is used (Schaible 1978, 1979). This technique has been frequently applied to combine the simple direct and the synthetic estimators (Schaible, Brock, and Schnack 1977, Royall 1973). However, deriving the optimal weight has generally been an insolvable problem in small area estimation. Although Schaible (1978, 1979) mentioned two conditions that might help in deriving the optimal weight, those conditions turned out to be hard to use. Schaible stated that the optimal weight could be found when the two components were independent and either of the two components was an unbiased estimator of the domain total. These assumptions are difficult to evaluate with respect to a sampling plan when one of the two components is the synthetic estimator and the other is the simple expansion estimator. The second condition Schaible mentioned that would allow the optimal weight to be found approximately was to assume that the covariance (with respect to the sampling plan) of the expansion and the synthetic estimators was small relative to the MSE of either of these components. Again, this condition is difficult to evaluate.

[1] Associate Professor, Department of Mathematical Sciences, College of Sciences, San Diego State University, San Diego, CA 92182–0314, U.S.A.
[2] Associate Professor, Department of Biostatistics, School of Public Health, University of California, Los Angeles, Los Angeles, CA 90024, U.S.A.

Inferences about finite populations can be made using approaches which do not require calculations with respect to the distribution created by the sampling plan. Lindley and Smith (1971) laid down a general framework for a Bayesian approach to estimation in finite population sampling. In the case of small area estimation, Purcell and Kish (1979) reviewed and discussed several techniques, indicating their performances and limitations. Fay and Herriot (1979) applied the James–Stein procedure to estimation of income for small areas, and recently, Fay (1987) discussed the use of components of variance models in small area estimation. A systematic collection of practical applications and theoretical developments in small area estimation appeared in Platek, Rao, Särndal, and Singh (1987). However, none of these papers discussed the estimation of optimal weights in composite estimation under a superpopulation model.

In this paper, we discuss the composite estimate of the uniformly minimum variance unbiased (UMVU) estimators under some random effects covariate models and derive the corresponding optimal weight in explicit form. We show how to estimate the optimal weight and suggest a test to help in deciding how to best apply the results to small domain estimation. Finally, a discussion of the composite of the simple direct estimator and the modified synthetic estimator is provided.

2. Composite Estimators

We suppose that the finite population is divided into I mutually exclusive subareas labeled $i = 1, \ldots, I$ for which we wish to produce estimates. Within each subdomain, units are further classified into J subgroups (for example, socioeconomic class, age, etc.); these are labeled $j = 1, \ldots, J$. The cell sizes N_{ij} resulting from this cross-classification are assumed to be known. Let y_{ijk} ($k = 1, 2, \ldots, N_{ij}$) be the measurement on the kth individual in the ijth cell and

$$T_i = \sum_{j=1}^{J} \sum_{k=1}^{N_{ij}} y_{ijk}$$

the total for the ith subdomain. The primary focus is to estimate the T_i's.

Letting s_{ij} denote the n_{ij} sampled units in the ijth cell, we use $\Sigma_{k \in s_{ij}} y_{ijk}$ to denote the sample sum and $\bar{y}_{ij.}$ to represent the average for the sampled units in cell ij. Similarly, let $\bar{y}_{.j.} = \Sigma_i \Sigma_{k \in s_{ij}} y_{ijk}/n_{.j}$, where $n_{.j} = \Sigma_i n_{ij}$.

The composite estimator of any two estimators $\hat{T}_i^{(1)}$ and $\hat{T}_i^{(2)}$ for T_i is defined as

$$\gamma \hat{T}_i^{(1)} + (1 - \gamma)\hat{T}_i^{(2)}, \quad \text{where} \quad 0 \leqslant \gamma \leqslant 1.$$

The optimal weight γ^* for the composite estimator to have the minimal MSE is given by

$$\gamma^* = [E(\hat{T}_i^{(2)} - T_i)^2 - E(\hat{T}_i^{(1)} - T_i) \times (\hat{T}_i^{(2)} - T_i)]/[E(\hat{T}_i^{(1)} - T_i)^2 + E(\hat{T}_i^{(2)} - T_i)^2 - 2E(\hat{T}_i^{(1)} - T_i)(\hat{T}_i^{(2)} - T_i)] \quad (1)$$

if $0 \leqslant \gamma^* \leqslant 1$, otherwise the optimal weight is equal to $\delta_{\{\gamma^* > 1\}}$, where δ is an indicator variable.

The composite estimator, $\gamma \hat{T}_i^{(1)} + (1 - \gamma)\hat{T}_i^{(2)}$, has MSE smaller than $\hat{T}_i^{(2)}$ if $\gamma < 2\gamma^*$. By symmetry, if $(1 - \gamma) < 2(1 - \gamma^*)$, then the composite estimator has MSE smaller than $\hat{T}_i^{(1)}$.

Traditionally, with respect to the distribution derived from the sampling plan, an attempt to find the optimal weight for the composite estimator is usually unsuccessful. This is because the formula for the optimal weight involves terms, such as MSE's of the estimators, that are difficult to evaluate with respect to the sampling plan. In the following discussion, in which we use a superpopulation model-based approach (Royall

1970), expectation is taken with respect to the distribution given in a model rather than the sampling plan.

3. Random Effects Covariate Models

Incorporating the implicit assumptions for the most common estimator, the synthetic estimator, for small area estimation, Holt, Smith, and Tomberlin (1979) derived the modified synthetic estimator. This estimator follows from the following model for the population structure

$$y_{ijk} = b_j + \varepsilon_{ijk} \tag{2}$$

where ε's are independent, normally distributed with mean 0 and variance σ^2. Furthermore, to incorporate knowledge from previous surveys to improve the precision of the modified synthetic estimator, Lui and Cumberland (1987, 1989) systematically extended the above model by assuming that the b_j are independent $N(\beta_0, \sigma_b^2)$ random variables, distributed independently of the ε's, and derived the generalized synthetic estimator. Knowledge obtained from a previous survey about the population under consideration can be represented in β_0. Discussions on the usefulness and applications of such models can be found elsewhere (Lui and Cumberland 1989). Note that if $\sigma_b^2 = 0$ or $\sigma_b^2 = \infty$, then we get one of the simple least-squares models that have been considered by Holt, Smith, and Tomberlin (1979). Because these random effects models are more flexible in borrowing information from other sources than the least-squares models, we derive the optimal weight of the composite estimator under this more general model. In the following discussion, we assume that the ratio of $\kappa = \sigma_b^2/\sigma^2$, which can be interpreted as the relative confidence of the prior knowledge to the current information, can be assigned by investigators, or is known. Methods to estimate this parameter can be found elsewhere (Ghosh and Meeden 1986, Dempster and Raghunathan 1986). When β_0 is known, the generalized synthetic estimator for T_i has been presented in our previous articles (Lui and Cumberland 1987, 1989) and it can be easily proved that this estimator under the random effects model always has a mean-squared-error smaller than the corresponding least-squares estimator. However, if β_0 is unknown, the corresponding generalized synthetic estimator does not necessarily outperform the corresponding least-squares estimator (Lui and Cumberland 1987, 1989). Scott and Smith (1969) used the super-population similar to that given in the above for estimation in multi-stage surveys, but did not discuss small area estimation.

In deriving the UMVU estimator of a subdomain total, we note that any finite subpopulation total can be represented as a sum of observed and unobserved random variables, so that the estimation problem becomes one of predicting the sum of the unobserved variables. Now, the UMVU estimator of this sum for unobserved variables is the sum of the conditional means, given the sampled units. Furthermore, we note that this sum of the condition means is a known linear function of parameters under our model assumptions (Lui and Cumberland 1987, 1989). Therefore, when the parameters are unknown, substitution of the UMVU estimators for these parameters will again give the UMVU estimator for the subdomain total (Graybill 1976). In other words, when β_0 is unknown, the UMVU estimator of T_i is given by

$$\hat{T}_i^{\text{GS1}} = \sum_j \sum_{k \in s_{ij}} y_{ijk} + \sum_j \sum_{k \notin s_{ij}} [(1 - \lambda_j)\bar{y}_w + \lambda_j \bar{y}_{.j.}] \tag{3}$$

where $\bar{y}_w = \Sigma_j \lambda_j \bar{y}_{.j.} / \Sigma_j \lambda_j \ (= \hat{\beta}_0)$ is the

UMVU estimator of β_0, and where $\lambda_j = n_{.j}\kappa/(n_{.j}\kappa + 1)$. The prediction variance of $\hat{T}_i^{GS1}$ is given by

$$V(\hat{T}_i^{GS1} - T_i) = \sum_j (N_{ij} - n_{ij})\sigma^2 + \sum_j (N_{ij} - n_{ij})^2 \lambda_j \sigma^2 / n_{.j} + \left(\sum_j (N_{ij} - n_{ij})(1 - \lambda_j)\right)^2 \sigma_b^2 \Big/ \left(\sum_j \lambda_j\right). \tag{4}$$

Often, we have some known covariate information, say x_j, that is related to the stratum effect b_j in the model (2). For example, y's may represent the number of physicians in the census tracts and x_j may represent the median house price, which could be used as a proxy for the measurement of socioeconomic class (Lui and Cumberland 1989). Other examples considering random effects covariate models may be found elsewhere (Platek et al. 1987).

In our model, we assume that b_j is normally distributed as N $(\beta_0 + \beta_1 x_j,\ \sigma_b^2)$. When β_0 and β_1 are unknown, we can apply similar arguments as before and show that the UMVU estimator of T_i under this model is given by

$$\hat{T}_i^{CS1} = \sum_j \sum_{k \in s_{ij}} y_{ijk} + \sum_j \sum_{k \notin s_{ij}} [(1 - \lambda_j)(\hat{\beta}_0 + \hat{\beta}_1 x_j) + \lambda_j \bar{y}_{.j.}] \tag{5}$$

where $\hat{\beta}_0 = \bar{y}_w - \hat{\beta}_1 \bar{x}_w$ and $\hat{\beta}_1 = \Sigma_j \lambda_j (x_j - \bar{x}_w)(\bar{y}_{.j.} - \bar{y}_w)/\Sigma_j \lambda_j (x_j - \bar{x}_w)^2$, and $\bar{x}_w = \Sigma_j \lambda_j x_j / \Sigma_j \lambda_j$. The prediction variance of $\hat{T}_i^{CS1}$ is given by

$$V(\hat{T}_i^{CS1} - T_i) = V(\hat{T}_i^{GS1} - T_i) + \left(\sum_j (N_{ij} - n_{ij})(1 - \lambda_j)(x_j - \bar{x}_w)\right)^2 \times \sigma_b^2 \Big/ \sum_j \lambda_j (x_j - \bar{x}_w)^2. \tag{6}$$

Note that the estimator $\hat{T}_i^{GS1}$, given by formula (3), though biased under the covariate model, can be a better estimator with respect to the MSE than the UMVU estimator $\hat{T}_i^{CS1}$ given by formula (5) under the random effects covariate model. This occurs when the coefficient of variation for $\hat{\beta}_1$ is large (Lui and Cumberland 1989). Determining the choice between $\hat{T}_i^{GS1}$ and $\hat{T}_i^{CS1}$ leads us to consider a weighted average of these two estimators that could have an MSE smaller than either $\hat{T}_i^{GS1}$ or $\hat{T}_i^{CS1}$. Using the facts that $E(\hat{T}_i^{CS1} - T_i)(\hat{T}_i^{GS1} - T_i) = V(\hat{T}_i^{GS1} - T_i)$ and $E(\hat{T}_i^{GS1} - T_i) = -\beta_1 \Sigma_j (N_{ij} - n_{ij})(1 - \lambda_j) \times (x_j - \bar{x}_w)$, we can use the variance formulae (4) and (6) in formula (1) to derive the optimal weight γ^* equal to $CV^2(\hat{\beta}_1)/(CV^2(\hat{\beta}_1) + 1)$, where $CV^2(\hat{\beta}_1) = V(\hat{\beta}_1)/(E(\hat{\beta}_1))^2$. Note that the condition $\gamma < 2\gamma^*$ is automatically satisfied if $CV^2(\hat{\beta}_1) > 1$. Therefore, $\gamma \hat{T}_i^{GS1} + (1 - \gamma)\hat{T}_i^{CS1}$ always has an MSE smaller than $\hat{T}_i^{CS1}$, if the coefficient of variation of $\hat{\beta}_1$ is greater than 1. Conversely, if $CV^2(\hat{\beta}_1) < 1$, then the composite estimator always has an MSE smaller than $\hat{T}_i^{GS1}$. Furthermore, we can easily show that the composite estimator always has an MSE smaller than either of its components if

a. $CV^2(\hat{\beta}_1) > 1$ and $\gamma \geqslant (CV^2(\hat{\beta}_1) - 1)/(CV^2(\hat{\beta}_1) + 1)$, or if

b. $CV^2(\hat{\beta}_1) < 1$ and $\gamma \leqslant 2CV^2(\hat{\beta}_1)/(CV^2(\hat{\beta}_1) + 1)$, or if

c. $CV^2(\hat{\beta}_1) = 1$.

Note that as $CV^2(\hat{\beta}_1) \to \infty$, the composite estimator with the optimal weight converges to $\hat{T}_i^{GS1}$. This implication is quite reasonable, because if the coefficient of variation $CV(\hat{\beta}_1)$ is very large, then using the information about β_1 in the estimator $\hat{T}_i^{CS1}$ might lead to a worse estimator than the estimator $\hat{T}_i^{GS1}$, which does not use this information.

Finally, the MSE of the composite estimator with the optimal weight γ^* is given by

$$E(\gamma^* \hat{T}_i^{GS1} + (1 - \gamma^*)\hat{T}_i^{CS1} - T_i)^2$$
$$= V(\hat{T}_i^{GS1} - T_i) + V(\hat{\beta}_1) \times [E(\hat{T}_i^{GS1} - T_i)]^2/[V(\hat{\beta}_1) + \beta_1^2]$$

where $V(\hat{T}_i^{GS1} - T_i)$, $V(\hat{\beta}_1)$, β_1, and the bias $(E(\hat{T}_i^{GS1} - T_i)$ can be easily estimated if κ, the relative size σ_b^2 to σ^2, is assumed to be known. Stroud (1987) and Dempster and Raghunathan (1987) discussed some other random effects covariate models similar to the one considered here, but did not discuss the composite estimator, nor did they consider the effect resulting from stratification in their papers.

We can generalize the above results to include any p-covariate ($p \leqslant J - 1$). Let $b_j \sim N(\beta_0 + \boldsymbol{\beta}'\mathbf{x}_j, \sigma_b^2)$, where $\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_p)$, and $\mathbf{x}_j' = (x_{j1}, x_{j2}, \ldots, x_{jp})$. In this p-covariate model, the UMVU estimator $\hat{T}_i^{CSP}$ of T_i is given by

$$\hat{T}_i^{CSP} = \sum_j \sum_{k \in s_{ij}} y_{ijk} + \sum_j \sum_{k \notin s_{ij}} [(1 - \lambda_j) \times (\hat{\beta}_0 + \hat{\beta}_1 x_{j1} + \cdots + \hat{\beta}_p x_{jp}) + \lambda_j \bar{y}_{.j.}]$$

where $\hat{\beta}_0 = \bar{y}_w - \Sigma_{t=1}^{P} \hat{\beta}_t \bar{x}_{wt}$ and $\hat{\boldsymbol{\beta}} = (\mathbf{X}_D' \boldsymbol{\Lambda}\, \mathbf{X}_D)^{-1} \mathbf{X}_D' \boldsymbol{\Lambda}\, \bar{\mathbf{y}}$

$$\mathbf{X}_D = (x_{jt} - \bar{x}_{wt})_{J \times P},$$
$$\bar{x}_{wt} = \sum_j \lambda_j x_{jt} \Big/ \sum_j \lambda_j$$
$$\boldsymbol{\Lambda} = (\text{diag}(\lambda_j))_{J \times J},$$
$$\bar{\mathbf{y}}' = (\bar{y}_{.1.}, \bar{y}_{.2.}, \ldots \bar{y}_{.J.}).$$

The prediction variance of $V(\hat{T}_i^{CSP} - T_i) = V(\hat{T}_i^{GS1} - T_i) + [\mathbf{L}_i'(\mathbf{X}_D'\boldsymbol{\Lambda}\mathbf{X}_D)^{-1}\mathbf{L}_i]\sigma_b^2$ where $\mathbf{L}_i' = (\Sigma_j(N_{ij} - n_{ij})(1 - \lambda_j)(x_{j1} - \bar{x}_{w1}), \ldots, \Sigma_j(N_{ij} - n_{ij})(1 - \lambda_j)(x_{jp} - \bar{x}_{wp}))$. In fact, every argument in the univariate case can be carried through simply replacing $CV^2(\beta_1)$ in the univariate case with $CV^2(\mathbf{L}_i'\boldsymbol{\beta})$.

4. Testing of $CV^2(\hat{\beta}_1)$

We have shown that $CV^2(\hat{\beta}_1) > 1$ ($CV^2(\hat{\beta}_1) < 1$) implies that the composite estimator $\gamma\hat{T}_i^{GS1} + (1 - \gamma)\hat{T}_i^{CS1}$ always has an MSE smaller than $\hat{T}_i^{CS1}(\hat{T}_i^{GS1})$. We must, however, decide whether $CV^2(\hat{\beta}_1) > 1$ or $CV^2(\hat{\beta}_1) < 1$, because usually the values of β_1 and σ_b^2 are unknown. One method for deciding whether $CV^2(\hat{\beta}_1) > 1$ is hypothesis testing, which requires that we find the distribution of the estimator of $CV^2(\hat{\beta}_1)$. We can easily rewrite the model assumption given in Section 3, with $b_j \sim N(\beta_0 + \beta_1 x_j, \sigma_b^2)$, into the matrix form

$$\begin{pmatrix} \mathbf{Y}_s \\ \mathbf{Y}_{\bar{s}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_s \\ \mathbf{X}_{\bar{s}} \end{pmatrix} \mathbf{B} + \begin{pmatrix} \boldsymbol{\varepsilon}_s \\ \boldsymbol{\varepsilon}_{\bar{s}} \end{pmatrix},$$

where $\binom{\boldsymbol{\varepsilon}_s}{\boldsymbol{\varepsilon}_{\bar{s}}}$ is distributed with $N(\mathbf{O}, \sigma^2\mathbf{I})$, $\mathbf{B} = (b_1, \ldots, b_J)'$ is distributed with $N(\mathbf{A}\boldsymbol{\beta}, \sigma_b^2\mathbf{I})$, and is independent of $\binom{\boldsymbol{\varepsilon}_s}{\boldsymbol{\varepsilon}_{\bar{s}}}$, $\mathbf{A}' = \binom{1, \; 1, \ldots \; 1}{x_1, x_2, \ldots x_J}_{2 \times J}$, $\boldsymbol{\beta}' = (\beta_0, \beta_1)$, and where s and $\bar{s}$ denote the sampled and non-sampled units, respectively.

Applying standard results for the general linear model (Graybill 1976), we get the UMVU estimator of σ^2

$$\hat{\sigma}_{CS1}^2 = \mathbf{Y}_s'[\mathbf{V}_s^{-1} - \mathbf{V}_s^{-1}\mathbf{X}_s^{(2)}(\mathbf{X}_s^{(2)'}\mathbf{V}_s^{-1}\mathbf{X}_s^{(2)})^{-1} \times \mathbf{X}_s^{(2)'}\mathbf{V}_s^{-1}]\mathbf{Y}_s/(n_{..} - 2)$$

where $\mathbf{Y}_s'$ is $1 \times n_{..}$ vector of measurements on the sampled individuals, $\mathbf{V}_s = \mathbf{I}_{n_{..} \times n_{..}} + \kappa\mathbf{X}_s\mathbf{X}_s'$, which is a block diagonal matrix, where the jth block is given by $\kappa \mathbf{1}n_{.j \times 1}\mathbf{1}_{1 \times n_{.j}}' + \mathbf{I}_{n_{.j} \times n_{.j}}$ and $\mathbf{X}_s^{(2)} = \mathbf{X}_s\mathbf{A}$. It is easy to show that $(n_{..} - 2)\hat{\sigma}_{CS1}^2/\sigma^2 \sim \chi^2(n_{..} - 2)$ which implies that $(n_{..} - 2)\kappa\hat{\sigma}_{CS1}/\sigma_b^2 \sim \chi^2(n_{..} - 2)$. Also, the UMVU estimator $\hat{\beta}_1$ of β_1 is

$$\hat{\beta}_1 = (0, 1)(\mathbf{X}_s^{(2)'}\mathbf{V}_s^{-1}\mathbf{X}_s^{(2)})^{-1}\mathbf{X}_s^{(2)'}\mathbf{V}_s^{-1}\mathbf{Y}_s$$

and $\hat{\beta}_1$ and $\hat{\sigma}_{CS1}^2$ are independent. Thus we have

$(\widehat{CV}^2(\hat{\beta}_1))^{-1} = \hat{\beta}_1^2(\hat{\sigma}_b^2/(\Sigma_j \lambda_j (x_j - \bar{x}_w)^2)^{-1}$, which is distributed as $F(1/[2CV^2(\beta_1)]; 1, n.. - 2)$, a noncentral F-distribution, where $\hat{\sigma}_b^2 = \kappa\hat{\sigma}_{CS1}^2$. We can use this resulting distribution to test the hypothesis $H_0 : CV^2(\hat{\beta}_1) = 1$ versus $(CV^2(\hat{\beta}_1))^{-1} > 1$. Under the null hypothesis, the distribution of the test statistic is an F-distribution with noncentrality parameter 0.5, leading to a simple test of the null hypothesis. When we reject the null hypothesis, we prefer using $\gamma\hat{T}_i^{GS1} + (1 - \gamma)\hat{T}_i^{CS1}$ (or in particular, $\hat{T}_i^{CS1}$) to using $\hat{T}_i^{GS1}$. This result is consistent with the fact that when $(CV^2(\hat{\beta}_1))^{-1}$ is very large, (implying that the prior knowledge about β_1 is very precise), the estimator $\hat{T}_i^{CS1}$ using the covariate information will be more accurate than the estimator $\hat{T}_i^{GS1}$ which ignores it. This test statistic can be easily generalized for the p-variate case. The test statistic for testing hypothesis $CV^2(\mathbf{L}_i'\hat{\beta}) = 1$ has a noncentral F-distribution with parameter equal to 0.5 and degrees of freedom 1 and $n.. - p - 1$.

5. Least-Squares Models Related to the Simple Direct and the Synthetic Estimators

Gonzalez and Waksberg (1973) compared the performances of the synthetic and the direct estimators. They concluded that when sample sizes in each small area were relatively small, the synthetic estimator outperformed the simple direct estimator, whereas, when sample sizes were relatively large, the simple direct outperformed the synthetic. These results suggested that a weighted sum of these two estimators would be a desirable alternative to choosing one over the other (Schaible 1978, Schaible, Brock, and Schnack 1977). Therefore, in this section, we focus on the estimator $\gamma\hat{T}_i^{D} + (1 - \gamma)\hat{T}_i^{MS}$, a linear combination of the simple direct estimator and the modified synthetic estimator (Holt, Smith, and Tomberlin 1979) under the super-population model-based approach.

The simple direct estimator

$$\hat{T}_i^{D} = \sum_j N_{ij}\bar{y}_{ij.}$$

is the UMVU estimator of T_i under the model

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

where ε's are independent, normally distributed with mean 0 and variance σ^2 and μ_{ij} are fixed unknown constants.

The modified synthetic estimator

$$\hat{T}_i^{MS} = \sum_j \sum_{k \in s_{ij}} y_{ijk} + \sum_j \sum_{k \notin s_{ij}} \bar{y}_{.j.}$$

is the UMVU estimator of T_i under the above model when $\mu_{1j} = \mu_{2j} = \ldots = \mu_{Ij}$. It is easy to show that $E(\hat{T}_i^{D} - T_i)(\hat{T}_i^{MS} - T_i) = V(\hat{T}_i^{MS} - T_i)$ under the non-restricted model. Therefore, we find the optimal weight from formula (1) is

$$\gamma^* = \left[\sum_j (N_{ij} - n_{ij})^2(1/n_{ij} - 1/n_{.j})\sigma^2\right] \Big/ \left[\sum_j (N_{ij} - n_{ij})^2(1/n_{ij} - 1/n_{.j})\sigma^2 + \left(\sum_j (N_{ij} - n_{ij})(\mu_{ij} - \bar{\mu}_{.j})\right)^2\right]$$

where $\bar{\mu}_{.j} = \Sigma_i n_{ij}\mu_{ij}/n_{.j}$. A reasonable estimate $\hat{\gamma}^*$ of γ^* can be obtained by substituting the UMVU estimators $\hat{\mu}_{ij} = \bar{y}_{ij.}$, $\hat{\bar{\mu}}_{.j} = \Sigma_i n_{ij}\bar{y}_{ij.}/n_{.j}$, and $\hat{\sigma}^2 = \Sigma_j\Sigma_i\Sigma_{k\in s_{ij}}(y_{ijk} - \bar{y}_{ij.})^2/(n.. - IJ)$ for μ_{ij}, $\bar{\mu}_{.j}$, and σ^2 in γ^* respectively.

A measure of error $E(\gamma\hat{T}_i^{D} + (1 - \gamma)\hat{T}_i^{MS} - T_i)^2$ for $\gamma\hat{T}_i^{D} + (1 - \gamma)\hat{T}_i^{MS}$ can be obtained from easily accessible estimates of $E(\hat{T}_i^{D} - T_i)^2$, $E(\hat{T}_i^{MS} - T_i)^2$, and $E(\hat{T}_i^{D} - T_i)(\hat{T}_i^{MS} - T_i)$.

6. Discussion

From the traditional point of view, Schaible (1979) pointed out two major problems in using the composite estimator. The first problem concerns how to estimate the optimal weight, given two estimators. The difficulty stems from the near impossibility of calculating, with respect to sampling plan, the MSE of small area estimators such as the synthetic estimator. Although Schaible (1979) suggested several different methods to estimate the optimal weight corresponding to different possible MSE's assumed for the population, these methods all depend on the true value T_i, and hence cannot be applied to estimate T_i in practice. The second problem, common to all small area estimators, is how to provide a measure of error of a composite estimator for a given small area. Using the model-based approach, however, we can estimate the optimal weight explicitly and provide the measure of error of the composite estimator for each small area. On the other hand, in the model-based approach, as proposed in this paper, we need to be cautious in evaluating the underlying model assumptions, on which the optimality of our estimators depends, although some robustness of our estimators has been published (Lui and Cumberland 1989). Using the sample data to estimate the optimal weight for random effects covariate models, as discussed here, it is generally not a problem in practice. However, when the expected cell mean is unknown, the estimated optimal weight for the composite estimator of the simple direct and the synthetic estimator depends on the cell sample means and these can be unreliable when the sample sizes in each cell are small. Using an unreliable estimate of the optimal weight in the composite estimator may not necessarily lead to an estimator which performs better than both of its components. Thus, in this situation, an empirical investigation of the performance of the composite estimator, similar to that given by Brock, French, and Peyton (1980), should be the next step in this research.

In summary, using the model-based approach, we have shown that the optimal weight of the composite estimator may be expressed explicitly in terms of the parameters in our model. Furthermore, we can also provide a measure of error for our estimators. Therefore, although one needs to be cautious when using the model-based estimators, the results given here should be useful to both survey statisticians and epidemiologists for estimating local area characteristics.

7. References

Brock, D.B., French, D.K., and Peyton, B.W. (1980). Small Area Estimation: Empirical Evaluation of Several Estimators for Primary Sampling Units. In Proceedings of the Section on Survey Research Methods, American Statistical Association, 766–771.

Dempster, A.P. and Raghunathan, T.E. (1986). A Method of Estimating the Interval Population of Counties. Journal of the American Statistical Association, 51, 587–590.

Dempster, A.P. and Raghunathan, T.E. (1987). Using a Covariate for Small Area Estimation: A Common Sense Bayesian Approach. In Small Area Statistics, An International Symposium, ed. R. Platek, J.N.K. Rao, C.E. Särndal, and M.P. Singh, New York: John Wiley & Sons, 77–90.

Fay, R.E. (1987). Small Domain Estimation Through Components of Variance Models. Bulletin of International Statistical Institute, 421–434.

Fay, R.E. and Herriot, R.A. (1979). Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. Journal of the American Statistical Association, 74, 269–277.

Ghosh, M. and Meeden, G. (1986). Empirical Bayes Estimation in Finite Population Sampling. Journal of the American Statistical Association, 81, 1058–1062.

Gonzalez, M.E. and Waksberg, J.E. (1973). Estimation of the Error of Synthetic Estimates. Presented at the first meeting of the International Association of Survey Statisticians, Vienna, Austria.

Graybill, F.A. (1976). Theory and Application of the Linear Model. North Scituate: Duxburg.

Holt, D., Smith, T.M.F., and Tomberlin, T.J. (1979). A Model-Based Approach to Estimation for Small Subgroups of a Population. Journal of the American Statistical Association, 74, 405–410.

Lindley, D.V. and Smith, A.F.M. (1971). Bayes Estimates for the Linear Model. Journal of Royal Statistical Society, ser. B, 34, 1–18.

Lui, K.-J. and Cumberland, W.G. (1987). A Bayesian Approach to Small Domain Estimation. In Proceedings of the Section on Survey Research Methods, American Statistical Association, 347–352.

Lui, K.-J. and Cumberland W.G. (1989). A Bayesian Approach to Small Domain Estimation. Journal of Official Statistics, 5, 143–156.

Platek, R., Rao, J.N.K., Särndal, C.E., and Singh, M.P. (1987). Small Area Statistics, An International Symposium. New York: John Wiley & Sons.

Purcell, N.J. and Kish, L. (1979). Estimation for Small Domains. Biometrics, 35, 365–384.

Royall, R.M. (1970). On Finite Population Sampling Theory under Certain Linear Regression Models. Biometrika, 57, 77–387.

Royall, R.M. (1973). Discussion of Two Papers on Recent Developments in Estimation of Local Areas. In Proceedings of the Section on Social Statistics, American Statistical Association, 43–44.

Schaible, W.L. (1978). Choosing Weight for Composite Estimators for Small Area Statistics. In Proceedings of the Section on Survey Research Methods, American Statistical Association, 741–746.

Schaible, W.L. (1979). A Composite Estimator for Small Area Statistics. In Synthetic Estimates for Small Areas: Statistical Workshop Papers and Discussion, ed. J. Steinberg, NIDA Research Monograph, 24, Rockville, Maryland, 36–59.

Schaible, W.L., Brock, D.B., and Schnack, G.A. (1977). An Empirical Comparison of the Simple Inflation, Synthetic and Composite Estimators for Small Area Statistics. In Proceedings of the Section on Social Statistics, American Statistical Association, 1017–1021.

Scott, A.J. and Smith, T.M.F. (1969). Estimation in Multi-stage Surveys. Journal of the American Statistical Association, 64, 830–840.

Stroud, T.W.F. (1987). Bayes and Empirical Bayes Approaches to Small Area Estimation. In Small Area Statistics, An International Symposium, ed. R. Platek, J.N.K. Rao, C.E. Särndal, and M.P. Singh, New York: John Wiley & Sons, 124–137.

Received February 1990
Revised September 1990