Small Area Estimation with Auxiliary Survey Data

by

Sharon L. Lohr [1] and N.G.N. Prasad [2]

[1]Department of Mathematics, Arizona State University, Tempe, AZ 85287-1804. Research partially supported by a grant from the U.S. Bureau of Justice Statistics.

[2]Department of Mathematical Sciences, University of Alberta, Edmonton, AB T6G 2G1. Research partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

Small Area Estimation with Auxiliary Survey Data

Abstract

Large governmental surveys typically provide accurate national statistics. To decrease the mean squared error of estimates for small areas—domains in which the sample size is small—often auxiliary variables from administrative records are used as covariates in a mixed linear model, and it is generally assumed that the auxiliary information is available for every small area in the population. In many cases, though, auxiliary information is available for only some of the small areas, either from another survey or from a previous administration of the same survey. We propose and develop properties of small area estimators that use multivariate models to combine information from several surveys. Computational algorithms are discussed, and a simulation study indicates that if quantities in the different surveys are sufficiently correlated, substantial gains in efficiency can be achieved.

Key Words and Phrases: best linear unbiased prediction; multiple surveys; multivariate mixed effects model; sampling on two occasions; variance components

1 Introduction

Most large sample surveys conducted by agencies such as the U.S. Bureau of the Census or by Statistics Canada provide accurate statistics at the national level. Increasingly, though, governments are interested in obtaining statistics for smaller domains such as states, provinces, or different racial and ethnic subgroups. These domains are called small areas—the term "small" refers to the fact that the sample size in the area or domain from the survey is small. The goal is to estimate μ_{iy}, the mean value of a variable of interest y in small area i.

National surveys such as the U.S. Current Population Survey (CPS) or the U.S. National Crime Victimization Survey (NCVS) are used to provide national estimates of poverty and criminal victimization. These surveys do not, however, contain sufficient sample sizes to give reliable estimates by themselves of small areas such as counties or minority groups, or to provide detailed information about events such as domestic violence that affect only a small part of the population. Current methods for estimating poverty in small areas incorporate auxiliary administrative information from sources such as tax records and food stamp programs as explanatory variables in a regression equation; the predicted value of the regression is combined with a direct estimate of poverty from the CPS to estimate the county poverty rate. This approach assumes that the administrative data are without errors; it also does not incorporate information from other surveys or make use of longitudinal information.

Traditionally, small area estimation relies on a mixed model relating the responses of interest in the small areas to each other and to covariates. The model allows the estimate of μ_{iy} to "borrow strength" from other small areas through random effects terms. Fay and Herriot (1979) first studied improved estimation in small areas using known vectors of covariate means. Since then, other models have been used by Dempster et al. (1981), Fuller and Harter (1987), and Battese et al. (1988), among others. Prasad and Rao (1990) put many of these estimators in a unified framework and derived second-order approximations to the mean squared errors of the estimators. Ghosh and Rao (1994), Marker (1999), and Rao (1999) reviewed much of the subsequent work in small area estimation. More recently, Datta et al. (1999) derived theory for multivariate small area estimation and Prasad and

Rao (1999) robustified estimation by incorporating the design weights.

In many situations, though, related information may be available for some units in a small area but not fit into one of the above frameworks. Another survey, with a possibly different sample design, may provide information related to the response of interest in the small areas. The new American Community Survey (ACS) is scheduled for full implementation in 2003; the National Research Council (1999) suggests using several years of ACS data to provide direct estimates of poverty, or using ACS data in a Fay-Herriot (1979) model. Since the CPS also provides information on poverty every year, however, greater precision in poverty estimates can be obtained by combining information from variables $\mathbf{x}$ measured from the ACS with variables y measured from the CPS.

For estimating criminal victimizations in small areas, the Uniform Crime Reporting System (UCR) provides data about crimes reported to the police in each state. The UCR data, however, are missing for many jurisdictions; some data points are likely misreported, and the types of crimes and definitions of crimes reported differ from those in the NCVS. Even though the UCR deals only with crimes reported to the police, measures of crimes from the UCR and NCVS are positively correlated for different metropolitan statistical areas (Wiersema et al., 2000). The UCR data can be used to improve the small-area estimates of victimization rates from the NCVS, even though the UCR has much missing and erroneous data; variables from the UCR can be thought of as values $\mathbf{x}$ from a second, independent survey much as the CPS and ACS would be considered separate surveys.

In another setting, many surveys have a panel design in which the same units may be sampled repeatedly. Households selected to participate in the Canadian Labour Force Survey remain in the sample for six consecutive months; each month, one-sixth of the sample is replaced with new households. As in Cochran's (1977, Section 12.11) description of sampling on two occasions, y is the value of a characteristic on the second occasion and the auxiliary variable x is the same variable for the first occasion.

In all of these settings, a primary response of interest, y, is measured in one survey; auxiliary responses $\mathbf{x}$ are measured in other surveys or in previous administration times of the same survey. We expect that y will be positively correlated with the auxiliary responses $\mathbf{x}$. Thus, we can expect an improvement in efficiency of small area estimates of μ_{iy} if the information from $\mathbf{x}$ is incorporated. However, since the information from $\mathbf{x}$ may have measurement error or sampling variability, using it as covariates in a regression model (as is done with administrative records) does not properly account for the uncertaintly in $\mathbf{x}$.

In this paper, we provide a general method for small area estimation when information is derived from two surveys or from repeated sampling of the same population. The multivariate approach adopted allows the variability in the auxiliary information $\mathbf{x}$ to be incorporated into the mean squared error of the estimates. In Section 2, we define the basic estimator and derive its mean squared error under the assumed multivariate mixed effects model. Section 3 gives a second-order asymptotic approximation to the MSE when fixed effects and covariance components are estimated from the data. Section 4 presents computational issues, and a simulation study is given in Section 5. Concluding comments are given in Section 6. All proofs are in the Appendix.

2 The model and estimator

2.1 Best Linear Unbiased Prediction

Suppose there are a total of t small areas; area i has N_i population units. Let y_{ij} denote the characteristic of interest for the j^{th} unit in area i, and let $\mathbf{x}_{ij} = (x_{ij1}, \cdots, x_{ijk})^T$ denote a vector of other characteristics for unit j of area i. For sampling on two occasions, $\mathbf{x}$ would represent measurements at time 1, and y would represent a measurement at time 2; with multiple surveys, $\mathbf{x}$ would be measured in one survey, and y in the other.

In area i, both $\mathbf{x}$ and y are measured on the n_i^{xy} units in $\mathcal{S}_{ixy}$; $\mathbf{x}$ (but not y) is measured on the n_i^x units in the set $\mathcal{S}_{ix}$; y (but not $\mathbf{x}$) is measured on the n_i^y units in the set $\mathcal{S}_{iy}$. If unit (ij) in the population is included in both samples, $m = k + 1$ measurements are recorded.

We use a multivariate mixed model to describe the relationship between $\mathbf{x}$, y, and covariates. In the following, $\mathbf{I}_j$ is the $j \times j$ identity matrix, $\mathbf{1}_j$ is a j-vector of ones, $\boldsymbol{\delta}_j$ is the m-vector with 1 in position j and zeroes elsewhere, $\bigoplus$ represents direct sum, and $\bigotimes$ represents Kronecker product. To simplify expression of results, we assume that the multivariate response vector $\mathbf{u}_i$ is arranged with all observations from $\mathcal{S}_{ixy}$ first, followed by those from $\mathcal{S}_{ix}$ and $\mathcal{S}_{iy}$, so

$$\mathbf{u}_i^T = [\mathbf{x}_{i1}^T, y_{i1}, \ldots, \mathbf{x}_{i,n_i^{xy}}^T, y_{i,n_i^{xy}}, \mathbf{x}_{i,n_i^{xy}+1}^T, \ldots, \mathbf{x}_{i,n_i^{xy}+n_i^x}^T, y_{i,n_i^{xy}+n_i^x+1}, \ldots, y_{i,n_i^{xy}+n_i^x+n_i^y}].$$

In the multivariate mixed model,

$$\mathbf{u} = \mathbf{A}\boldsymbol{\mu} + \mathbf{Z}\mathbf{v} + \mathbf{e} \tag{1}$$

where $\mathbf{u}^T = (\mathbf{u}_1^T \cdots \mathbf{u}_t^T)$, $\boldsymbol{\mu}$ is a vector of fixed effects parameters, $\mathbf{A}$ and $\mathbf{Z}$ are known matrices, and $\mathbf{v} = (\mathbf{v}_1^T \cdots \mathbf{v}_t^T)$ and $\mathbf{e}$ are independent random vectors with mean $\mathbf{0}$ and respective covariance matrices $\mathbf{G}$ and $\mathbf{R}$. We assume that observations in different small areas are independent so that $\mathbf{R} = \mathrm{Cov}\,(\mathbf{e}) = \bigoplus_{i=1}^t \mathbf{R}_i$ and $\mathbf{G} = \mathrm{Cov}\,(\mathbf{v}) = \bigoplus_{i=1}^t \mathbf{G}_i$. The overall covariance matrix of $\mathbf{u}$ is

$$\mathbf{V} = \mathrm{Cov}\,(\mathbf{u}) = \mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}^T.$$

Since we are interested in estimating means for all small areas, including those for which either $\mathbf{x}$ or y is not measured, we take $\mathbf{v}_i$ to be a random m-vector of all the random effects for area i, even if either $\mathbf{x}$ or y is not measured in the area. Under this representation, $\mathbf{Z} = \bigoplus_{i=1}^t \mathbf{Z}_i$, where

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{1}_{n_i^{xy}} & \otimes & \mathbf{I}_m \\ \mathbf{1}_{n_i^x} & \otimes & (\mathbf{I}_k \ \mathbf{0}_k) \\ \mathbf{1}_{n_i^y} & \otimes & (\mathbf{0}_k^T \ 1) \end{bmatrix}$$

For simplicity of presentation, we take $\boldsymbol{\mu}$ to be the m-vector of fixed effects means, partitioned as $\boldsymbol{\mu}^T = [\boldsymbol{\mu}_x^T \ \mu_y]$. Then $\mathbf{A}^T = [\mathbf{Z}_1^T \, \mathbf{Z}_2^T \ \ldots \ \mathbf{Z}_t^T]$. However, all results in this paper are easily extended to the case where $\boldsymbol{\mu}$ is a general vector of parameters, and $\mathbf{A}$ is a matrix of fixed effects covariates. In this way information from a census or from administrative records may be incorporated into the small area estimates through regression.

We assume $\mathbf{G}_i = \mathrm{Cov}\,(\mathbf{v}_i) = \boldsymbol{\Sigma}_v$ for all i and that

$$\mathbf{R}_i = [\mathbf{I}_{n_i^{xy}} \bigotimes \boldsymbol{\Sigma}_e] \bigoplus [\mathbf{I}_{n_i^x} \bigotimes \boldsymbol{\Sigma}_{exx}] \bigoplus [\mathbf{I}_{n_i^y} \bigotimes \boldsymbol{\Sigma}_{eyy}],$$

where the matrices $\boldsymbol{\Sigma}_v$ and $\boldsymbol{\Sigma}_e$ are partitioned as

$$\boldsymbol{\Sigma}_v = \begin{bmatrix} \boldsymbol{\Sigma}_{vxx} & \boldsymbol{\Sigma}_{vxy} \\ \boldsymbol{\Sigma}_{vxy}^T & \boldsymbol{\Sigma}_{vyy} \end{bmatrix}, \quad \boldsymbol{\Sigma}_e = \begin{bmatrix} \boldsymbol{\Sigma}_{exx} & \boldsymbol{\Sigma}_{exy} \\ \boldsymbol{\Sigma}_{exy}^T & \boldsymbol{\Sigma}_{eyy} \end{bmatrix}.$$

With this assumption of equal variances for the small area means, $\mathbf{G} = \mathrm{Cov}\,(\mathbf{v}) = \mathbf{I}_t \bigotimes \mathbf{\Sigma}_v$. Thus $\mathbf{V} = \bigoplus_{i=1}^t \mathbf{V}_i$, where

$$\mathbf{V}_i = \mathrm{Cov}\,(\mathbf{u}_i) = \mathbf{R}_i + \mathbf{Z}_i \mathbf{\Sigma}_v \mathbf{Z}_i^T. \tag{2}$$

The matrices $\mathbf{\Sigma}_v$ and $\mathbf{\Sigma}_e$ are assumed positive definite, and may be written as functions of a vector of variances and covariances denoted by $\boldsymbol{\theta}$. Note that if there is no auxiliary information (that is, $n_i^{xy} = n_i^x = 0$), the model and assumption reduce to those in Battese et al. (1988).

Under this model, the vector of means for small area i is $\boldsymbol{\mu}_i = \boldsymbol{\mu} + \mathbf{v}_i$. Theorem 1 below gives the best linear unbiased predictor (BLUP) of $\boldsymbol{\mu}_i$ under the model in (1). In the following, let

$$\mathbf{\Sigma}_e^* = \begin{bmatrix} \mathbf{\Sigma}_{exx} & 0 \\ 0 & \mathbf{\Sigma}_{eyy} \end{bmatrix}, \tag{3}$$

$$\mathbf{n}_i^* = \begin{bmatrix} n_i^x \mathbf{I} & 0 \\ 0 & n_i^y \end{bmatrix}, \tag{4}$$

$$\mathbf{E}_i = \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i = n_i^{xy} \mathbf{\Sigma}_e^{-1} + \mathbf{n}_i^* (\mathbf{\Sigma}_e^*)^{-1}, \tag{5}$$

and

$$\mathbf{D}_i = (\mathbf{\Sigma}_v^{-1} + \mathbf{E}_i)^{-1}. \tag{6}$$

Also, for n_i^{xy}, n_i^x, or n_i^y nonzero, define

$$\bar{\mathbf{u}}_{ixy} = \frac{1}{n_i^{xy}} \sum_{j=1}^{n_i^{xy}} \begin{pmatrix} \mathbf{x}_{ij} \\ y_{ij} \end{pmatrix}, \quad \bar{\mathbf{x}}_{ix} = \frac{1}{n_i^x} \sum_{j=n_i^{xy}+1}^{n_i^{xy}+n_i^x} \mathbf{x}_{ij}, \quad \bar{y}_{iy} = \frac{1}{n_i^y} \sum_{j=n_i^{xy}+n_i^x+1}^{n_i^{xy}+n_i^x+n_i^y} y_{ij}, \quad \bar{\mathbf{u}}_i^* = \begin{pmatrix} \bar{\mathbf{x}}_{ix} \\ \bar{y}_{iy} \end{pmatrix}$$

Theorem 1 *Assume $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$ are known, and that $\mathbf{\Sigma}_v$ and $\mathbf{\Sigma}_e$ are positive definite. Then the BLUP of $\boldsymbol{\mu}_i$ under the multivariate mixed model in (1) is*

$$\tilde{\boldsymbol{\mu}}_i = \boldsymbol{\mu} + n_i^{xy} \mathbf{D}_i \mathbf{\Sigma}_e^{-1} (\bar{\mathbf{u}}_{ixy} - \boldsymbol{\mu}) + \mathbf{D}_i \mathbf{n}_i^* (\mathbf{\Sigma}_e^*)^{-1} (\bar{\mathbf{u}}_i^* - \boldsymbol{\mu}). \tag{7}$$

In addition,

$$MSE[\tilde{\boldsymbol{\mu}}_i] = \mathbf{D}_i. \tag{8}$$

Using (7), the small area estimate for the mean of response y in area i is $\tilde{\boldsymbol{\mu}}_{iy} = \boldsymbol{\delta}_m^T \tilde{\boldsymbol{\mu}}_i$. The mean squared error of $\tilde{\boldsymbol{\mu}}_{iy}$ is $\mathbf{D}_{iyy}$, the (m, m) entry of $\mathbf{D}_i$.

Two special cases of Theorem 1 are of interest. First, suppose that both $\mathbf{x}$ and y are measured on every sampled unit in area i, so that $n_i^x = n_i^y = 0$. This would happen, for example, if $\mathbf{x}$ is the characteristic of interest at time 1, y is the same characteristic at time 2, and the same units are included in the sample for each occasion.

Corollary 1 *Suppose the conditions of Theorem 1 hold, and that* $n_i^x = n_i^y = 0$. *Then*

$$\tilde{\boldsymbol{\mu}}_i = \boldsymbol{\mu} + \mathbf{D}_i \mathbf{E}_i (\bar{\mathbf{u}}_{ixy} - \boldsymbol{\mu})$$

In this special case, our estimator reduces to the multivariate BLUP given in Datta et al. (1999).

In the second special case, suppose that y is not measured in area i, so that estimation of μ_{iy} depends entirely on the auxiliary information.

Corollary 2 *Suppose the conditions of Theorem 1 hold, and that* $n_i^{xy} = n_i^y = 0$. *Let* $\mathbf{E}_i^x = n_i^x \boldsymbol{\Sigma}_{exx}^{-1}$ *and* $\mathbf{D}_i^x = (\boldsymbol{\Sigma}_{vxx}^{-1} + \mathbf{E}_i^x)^{-1}$. *Then*

$$\mathbf{D}_i = \boldsymbol{\Sigma}_v \begin{bmatrix} \boldsymbol{\Sigma}_{vxx}^{-1} \mathbf{D}_i^x & -\boldsymbol{\Sigma}_{vxx}^{-1} \mathbf{D}_i^x \mathbf{E}_i^x \boldsymbol{\Sigma}_{vxy} \\ 0 & 1 \end{bmatrix}$$

and

$$\tilde{\boldsymbol{\mu}}_i = \boldsymbol{\mu} + \begin{pmatrix} \mathbf{I}_k \\ \boldsymbol{\Sigma}_{vxy}^T \boldsymbol{\Sigma}_{vxx}^{-1} \end{pmatrix} \mathbf{D}_i^x \mathbf{E}_i^x (\bar{\mathbf{x}}_{ix} - \boldsymbol{\mu}_x).$$

When $n_i^{xy} = n_i^y = 0$, $\boldsymbol{\mu}_{ix}$ is estimated by the multivariate BLUP $\tilde{\boldsymbol{\mu}}_{ix} = \boldsymbol{\mu}_x + \mathbf{D}_i^x \mathbf{E}_i^x (\bar{\mathbf{x}}_{ix} - \boldsymbol{\mu}_x)$ from Corollary 1, calculated assuming only data from the $\mathbf{x}$'s are present. The BLUP of μ_{iy} when $n_i^{xy} = n_i^y = 0$ is $\tilde{\mu}_{iy} = \mu_y + \boldsymbol{\Sigma}_{vxy}^T \boldsymbol{\Sigma}_{vxx}^{-1} (\tilde{\boldsymbol{\mu}}_{ix} - \boldsymbol{\mu}_x)$.

2.2 Relative Efficiency

We next consider the efficiency of $\tilde{\mu}_{iy}$ relative to the estimator that does not use auxiliary information from another survey. If the information in $\mathbf{x}$ is not used, the univariate small area estimate of μ_{iy} under the model in (1) is

$$\tilde{\mu}_{iy}^{univ} = \mu_y + [\boldsymbol{\Sigma}_{vyy}^{-1} + (n_i^{xy} + n_i^y) \boldsymbol{\Sigma}_{eyy}^{-1}]^{-1} \boldsymbol{\Sigma}_{eyy}^{-1} \{n_i^{xy} [\bar{\mathbf{u}}_{ixy}]_m + n_i^y \bar{y}_{iy} - (n_i^{xy} + n_i^y) \mu_y\}$$

with mean squared error

$$\text{MSE}\,[\tilde{\mu}_{iy}^{univ}] = [\mathbf{\Sigma}_{vyy}^{-1} + (n_i^{xy} + n_i^{y})\mathbf{\Sigma}_{eyy}^{-1}]^{-1} = \frac{\mathbf{\Sigma}_{vyy}\mathbf{\Sigma}_{eyy}}{\mathbf{\Sigma}_{eyy} + (n_i^{xy} + n_i^{y})\mathbf{\Sigma}_{vyy}}.$$

Let $\rho_v^2 = \mathbf{\Sigma}_{vxy}^T \mathbf{\Sigma}_{vxx}^{-1} \mathbf{\Sigma}_{vxy} \mathbf{\Sigma}_{vyy}^{-1}$ and $\rho_e^2 = \mathbf{\Sigma}_{exy}^T \mathbf{\Sigma}_{exx}^{-1} \mathbf{\Sigma}_{exy} \mathbf{\Sigma}_{eyy}^{-1}$. Also define $\mathbf{c} = \mathbf{\Sigma}_{vyy}^{-1} \mathbf{\Sigma}_{vxx}^{-1} \mathbf{\Sigma}_{vxy}$, $\mathbf{d} = \mathbf{\Sigma}_{eyy}^{-1} \mathbf{\Sigma}_{exx}^{-1} \mathbf{\Sigma}_{exy}$, $\mathbf{h} = (1-\rho_v^2)^{-1}\mathbf{c} + n_i^{xy}(1-\rho_e^2)^{-1}\mathbf{d}$, and

$$\mathbf{M} = \mathbf{\Sigma}_{vxx}^{-1} + (1-\rho_v^2)^{-1}\mathbf{c}\mathbf{\Sigma}_{vyy}\mathbf{c}^T + (n_i^{xy} + n_i^{x})\mathbf{\Sigma}_{exx}^{-1} + n_i^{xy}(1-\rho_e^2)^{-1}\mathbf{d}\mathbf{\Sigma}_{eyy}\mathbf{d}^T.$$

Theorem 2 gives a simplified expression for $\text{MSE}\,[\tilde{\boldsymbol{\mu}}_{iy}] = \mathbf{D}_{iyy}$.

Theorem 2 *Suppose the conditions of Theorem 1 hold, and that $\rho_v^2 < 1$ and $\rho_e^2 < 1$. Then*

$$\mathbf{D}_{iyy} = [(1-\rho_v^2)^{-1}\mathbf{\Sigma}_{vyy}^{-1} + n_i^{y}\mathbf{\Sigma}_{eyy}^{-1} + n_i^{xy}(1-\rho_e^2)^{-1}\mathbf{\Sigma}_{eyy}^{-1} - \mathbf{h}^T\mathbf{M}^{-1}\mathbf{h}]^{-1}. \tag{9}$$

The multivariate estimator is always at least as efficient as the univariate estimator; the gain in efficiency is given in the following corollary.

Corollary 3 *Suppose that the conditions of Theorem 2 hold. Let $\mathbf{Q}$ be an orthogonal matrix and $\mathbf{\Lambda}$ a diagonal matrix such that $\mathbf{\Sigma}_{vxx}^{-1/2}\mathbf{\Sigma}_{exx}\mathbf{\Sigma}_{vxx}^{-1/2} = \mathbf{Q}^T\mathbf{\Lambda}\mathbf{Q}$, and let $\mathbf{P} = \mathbf{Q}\mathbf{\Sigma}_{vxx}^{1/2}$. Then*

$$\mathbf{D}_{iyy} = MSE\,[\tilde{\mu}_{iy}^{univ}]\left\{1 - \mathbf{D}_{iyy}\left[\mathbf{b}^T\mathbf{C}^{-1}\mathbf{b} + \mathbf{\Sigma}_{eyy}^{-1}[n_i^{xy} + n_i^{x}(1-\rho_e^2)]^{-1}n_i^{x}n_i^{xy}\rho_e^2\right]\right\}, \tag{10}$$

where

$$\mathbf{b} = \mathbf{P}\mathbf{c} - [n_i^{xy} + n_i^{x}(1-\rho_e^2)]^{-1}n_i^{xy}\mathbf{\Lambda}\mathbf{P}\mathbf{d}$$

and where

$$\mathbf{C} = \mathbf{I} - \mathbf{\Sigma}_{vyy}\mathbf{P}\mathbf{c}\mathbf{c}^T\mathbf{P}^T + (n_i^{xy} + n_i^{x})^{-1}\left[\mathbf{\Lambda} - [n_i^{xy} + n_i^{x}(1-\rho_e^2)]^{-1}n_i^{xy}\mathbf{\Sigma}_{eyy}\mathbf{\Lambda}\mathbf{P}\mathbf{d}\mathbf{d}^T\mathbf{P}^T\mathbf{\Lambda}\right]$$

is positive definite.

The estimator $\tilde{\mu}_{iy}$ is thus always more efficient than $\tilde{\mu}_{iy}^{univ}$ if $n_i^{x}(\rho_v^2 + n_i^{xy}\rho_e^2) > 0$ or if $\mathbf{\Sigma}_{eyy}\mathbf{\Sigma}_{vxy} \neq \mathbf{\Sigma}_{vyy}\mathbf{\Sigma}_{exy}$. The mean squared errors of the two estimators are equivalent if $\mathbf{\Sigma}_v = \mathbf{\Sigma}_e$ and $n_i^{x} = n_i^{y} = 0$.

The expression in Corollary 3 simplifies when $m = 2$: then,

$$\begin{aligned}\mathbf{D}_{iyy} &= \text{MSE}\,[\tilde{\mu}_{iy}^{univ}]\left\{1 - \frac{\mathbf{D}_{iyy}}{\mathbf{M}(1-\rho_v^2)(1-\rho_e^2)}\left[n_i^{xy}\left(\frac{\rho_v}{\sqrt{\mathbf{\Sigma}_{vyy}\mathbf{\Sigma}_{exx}}} - \frac{\rho_e}{\sqrt{\mathbf{\Sigma}_{vxx}\mathbf{\Sigma}_{eyy}}}\right)^2\right.\right. \\ &\qquad \left.\left. + \; n_i^{x}\frac{\rho_v^2(1-\rho_e^2)}{\mathbf{\Sigma}_{vyy}\mathbf{\Sigma}_{exx}} + n_i^{x}n_i^{xy}\frac{\rho_e^2(1-\rho_v^2)}{\mathbf{\Sigma}_{exx}\mathbf{\Sigma}_{eyy}}\right]\right\}.\end{aligned}$$

3 Estimation when variance components are unknown

In the results of Section 2, we assumed that $\boldsymbol{\mu}$ and all multivariate variance components were known. In practice, these may need to be estimated from the data. To find a second-order approximation to the MSE of the estimators when the quantities are estimated, we first examine the case when only $\boldsymbol{\mu}$ is unknown. Then the generalized least squares estimator of $\boldsymbol{\mu}$ is

$$\tilde{\boldsymbol{\mu}} = (\mathbf{A}^T\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{V}^{-1}\mathbf{u}. \tag{11}$$

Then, letting

$$\hat{\tilde{\boldsymbol{\mu}}}_i = \tilde{\boldsymbol{\mu}} + n_i^{xy}\mathbf{D}_i\boldsymbol{\Sigma}_e^{-1}(\bar{\mathbf{u}}_{ixy} - \tilde{\boldsymbol{\mu}}) + \mathbf{D}_i\mathbf{n}_i^*(\boldsymbol{\Sigma}_e^*)^{-1}(\bar{\mathbf{u}}_i^* - \tilde{\boldsymbol{\mu}}), \tag{12}$$

it may be shown directly that

$$\text{MSE}\,[\hat{\tilde{\boldsymbol{\mu}}}_i] = \mathbf{D}_i + (\mathbf{I} - \mathbf{D}_i\mathbf{E}_i)(\mathbf{A}^T\mathbf{V}^{-1}\mathbf{A})^{-1}(\mathbf{I} - \mathbf{E}_i\mathbf{D}_i). \tag{13}$$

The estimator in (12), however, still depends on the vector of variance components $\boldsymbol{\theta}$, which may be unknown. Let $\hat{\boldsymbol{\theta}}$ be a consistent estimator of $\boldsymbol{\theta}$. Then we may estimate $\boldsymbol{\mu}_i$ by

$$\hat{\boldsymbol{\mu}}_i = \hat{\boldsymbol{\mu}} + n_i^{xy}\hat{\mathbf{D}}_i\hat{\boldsymbol{\Sigma}}_e^{-1}(\bar{\mathbf{u}}_{ixy} - \hat{\boldsymbol{\mu}}) + \hat{\mathbf{D}}_i\mathbf{n}_i^*(\hat{\boldsymbol{\Sigma}}_e^*)^{-1}(\bar{\mathbf{u}}_i^* - \hat{\boldsymbol{\mu}}), \tag{14}$$

where

$$\hat{\boldsymbol{\mu}} = (\mathbf{A}^T\hat{\mathbf{V}}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\hat{\mathbf{V}}^{-1}\mathbf{u}$$

and the estimators $\hat{\mathbf{D}}_i$, $\hat{\boldsymbol{\Sigma}}_e$, $\hat{\boldsymbol{\Sigma}}_e^*$, and $\hat{\mathbf{V}}$ are formed by substituting $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ in the corresponding quantities $\mathbf{D}_i, \boldsymbol{\Sigma}_e$, $\boldsymbol{\Sigma}_e^*$ and $\mathbf{V}$.

Theorem 3 gives the second-order asymptotic MSE of $\hat{\boldsymbol{\mu}}_i$ when method of moment estimators are used to estimate components of $\boldsymbol{\Sigma}_e$ and $\boldsymbol{\Sigma}_v$. Datta and Lahiri (2000) discuss the second-order MSE for the univariate case when maximum likelihood or restricted maximum likelihood estimates are used; their arguments can be extended to the multivariate situation in similar fashion.

Theorem 3 *Assume $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = O_p(t^{-1})$ and that the regularity conditions in Section A.1 of Prasad and Rao (1990) hold. Then the mean squared error of the estimator in (14) is*

$$MSE[\hat{\boldsymbol{\mu}}_i] = MSE[\hat{\tilde{\boldsymbol{\mu}}}_i] + \mathbf{T}_i + o(t^{-1}), \tag{15}$$

where $\mathbf{T}_i$ *is an* $m \times m$ *matrix whose* (j,l) *element is* $tr\{\mathbf{G}_{jl}E[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T]\}$, *and the* (q,r) *element of* $\mathbf{G}_{jl}$ *is*

$$\begin{aligned}[\mathbf{G}_{jl}]_{qr} &= \boldsymbol{\delta}_j^T \left\{ n_i^{xy} \frac{\partial \mathbf{D}_i \boldsymbol{\Sigma}_e^{-1}}{\partial \boldsymbol{\theta}_q} \boldsymbol{\Sigma}_e \left(\frac{\partial \mathbf{D}_i \boldsymbol{\Sigma}_e^{-1}}{\partial \boldsymbol{\theta}_r} \right)^T + \frac{\partial \mathbf{D}_i (\boldsymbol{\Sigma}_e^*)^{-1}}{\partial \boldsymbol{\theta}_q} \mathbf{n}_i^* \boldsymbol{\Sigma}_e^* \left(\frac{\partial \mathbf{D}_i (\boldsymbol{\Sigma}_e^*)^{-1}}{\partial \boldsymbol{\theta}_r} \right)^T \right. \\ &\qquad \left. + \frac{\partial (\mathbf{D}_i \mathbf{E}_i)}{\partial \boldsymbol{\theta}_q} \boldsymbol{\Sigma}_v \left(\frac{\partial (\mathbf{D}_i \mathbf{E}_i)}{\partial \boldsymbol{\theta}_r} \right)^T \right\} \boldsymbol{\delta}_l. \end{aligned} \tag{16}$$

In particular, let $\boldsymbol{\theta}$ *be the vector consisting of the distinct entries of* $\boldsymbol{\Sigma}_v$ *and* $\boldsymbol{\Sigma}_e$. *Define* $\boldsymbol{\Delta}_{ab}$ *to be the* $m \times m$ *matrix with ones in positions* (a,b) *and* (b,a) *and zeroes elsewhere. Also define* η_{ab} *to be 1 if* $a \in \{1,\ldots,k\}$ *and* $b \in \{1,\ldots,k\}$, *1 if* $a = m$ *and* $b = m$, *and 0 otherwise. Then for* $a,b,c,d \in \{1,\ldots,m\}$,

$$[\mathbf{G}_{jl}]_{qr} = \begin{cases} \boldsymbol{\delta}_j^T \mathbf{D}_i \boldsymbol{\Sigma}_v^{-1} \boldsymbol{\Delta}_{ab} (\mathbf{I} - \boldsymbol{\Sigma}_v^{-1} \mathbf{D}_i) \boldsymbol{\Sigma}_v^{-1} \boldsymbol{\Delta}_{cd} \boldsymbol{\Sigma}_v^{-1} \mathbf{D}_i \boldsymbol{\delta}_l, \ \text{if } \boldsymbol{\theta}_q = [\boldsymbol{\Sigma}_v]_{ab}, \boldsymbol{\theta}_r = [\boldsymbol{\Sigma}_v]_{cd} \\ \boldsymbol{\delta}_j^T \mathbf{D}_i \left\{ n_i^{xy} \boldsymbol{\Sigma}_e^{-1} \boldsymbol{\Delta}_{ab} \boldsymbol{\Sigma}_e^{-1} \boldsymbol{\Delta}_{cd} \boldsymbol{\Sigma}_e^{-1} + \eta_{ab}\eta_{cd} (\boldsymbol{\Sigma}_e^*)^{-1} \boldsymbol{\Delta}_{ab} \mathbf{n}_i^* (\boldsymbol{\Sigma}_e^*)^{-1} \boldsymbol{\Delta}_{cd} (\boldsymbol{\Sigma}_e^*)^{-1} \right. \\ \qquad \left. - \mathbf{Q}_{iab} \mathbf{D}_i \mathbf{Q}_{icd} \right\} \mathbf{D}_i \boldsymbol{\delta}_l, \ \text{if } \boldsymbol{\theta}_q = [\boldsymbol{\Sigma}_e]_{ab}, \boldsymbol{\theta}_r = [\boldsymbol{\Sigma}_e]_{cd} \\ -\boldsymbol{\delta}_j^T \mathbf{D}_i \mathbf{Q}_{iab} \mathbf{D}_i \boldsymbol{\Sigma}_v^{-1} \boldsymbol{\Delta}_{cd} \boldsymbol{\Sigma}_v^{-1} \mathbf{D}_i \boldsymbol{\delta}_l, \ \text{if } \boldsymbol{\theta}_q = [\boldsymbol{\Sigma}_e]_{ab}, \boldsymbol{\theta}_r = [\boldsymbol{\Sigma}_v]_{cd} \end{cases} \tag{17}$$

where $\mathbf{Q}_{iab} = n_i^{xy} \boldsymbol{\Sigma}_e^{-1} \boldsymbol{\Delta}_{ab} \boldsymbol{\Sigma}_e^{-1} + \mathbf{n}_i^* (\boldsymbol{\Sigma}_e^*)^{-1} \boldsymbol{\Delta}_{ab} (\boldsymbol{\Sigma}_e^*)^{-1} \eta_{ab}$.

Equations (14) and (17) require $\boldsymbol{\Sigma}_{exy}$ to be estimated only when $n_i^{xy} > 0$. For small areas with $n_i^{xy} > 0$, the assumption $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = O_p(t^{-1})$ requires that n_i^{xy} must be positive in a sufficient number of the small areas to allow $\boldsymbol{\Sigma}_e$ to be estimated with sufficient precision.

4 Calculating estimates

Because of the complexity caused by the multivariate structure and the possibility of unbalanced data, many computational challenges arise when calculating estimates of $\tilde{\boldsymbol{\mu}}_i$.

First consider the case when $\boldsymbol{\Sigma}_e$ and $\boldsymbol{\Sigma}_v$ are known. We do not recommend using the formulas in (7) or (14) for estimation because they may be numerically unstable and computationally inefficient. If $\boldsymbol{\Sigma}_v$ is close to being singular, then the formula in (6) for $\mathbf{D}_i$ may lead to inaccuracies in computation. Instead, write $\boldsymbol{\Sigma}_v = \mathbf{U}\mathbf{U}^T$, where $\mathbf{U}$ is upper triangular. Then, writing $\mathbf{D}_i$ as $\mathbf{D}_i = \mathbf{U}[\mathbf{I} + \mathbf{U}^T \mathbf{E}_i \mathbf{U}]^{-1} \mathbf{U}^T$ does not require inversion of $\boldsymbol{\Sigma}_v$.

If $n_i^x = n_i^y = 0$ for each i, the computations may be further simplified. In that case, write $\mathbf{U}^T \boldsymbol{\Sigma}_e^{-1} \mathbf{U} = \mathbf{P} \boldsymbol{\Lambda} \mathbf{P}^T$ where $\mathbf{P}$ is orthogonal and $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues. Then $\mathbf{D}_i = \mathbf{U}\mathbf{P}[\mathbf{I} + n_i^{xy}\boldsymbol{\Lambda}]^{-1}\mathbf{P}^T\mathbf{U}^T$; this form reduces the computations from $t+2$ matrix inversions to one inversion, one Cholesky decomposition, and one spectral decomposition.

Computing $\mathbf{D}_i$ is more involved for the general case, but the following may be used to improve numerical stability. Write

$$\begin{bmatrix} \boldsymbol{\Sigma}_e & 0 \\ 0 & \boldsymbol{\Sigma}_e^* \end{bmatrix} = \mathbf{T}\mathbf{T}^T,$$

for $\mathbf{T}$ upper triangular. Then

$$\mathbf{E}_i = \begin{bmatrix} \sqrt{n_i^{xy}}\, \mathbf{I}_m & \sqrt{\mathbf{n}_i^*} \end{bmatrix} \mathbf{T}^{-T}\mathbf{T}^{-1} \begin{bmatrix} \sqrt{n_i^{xy}}\, \mathbf{I}_m \\ \sqrt{\mathbf{n}_i^*} \end{bmatrix}$$

and

$$\mathbf{D}_i = \mathbf{U} \left\{ \left[\mathbf{U}^T \begin{pmatrix} \sqrt{n_i^{xy}}\, \mathbf{I}_m & \sqrt{\mathbf{n}_i^*} \end{pmatrix} \mathbf{T}^{-T},\ \mathbf{I} \right] \begin{bmatrix} \mathbf{T}^{-1} \begin{pmatrix} \sqrt{n_i^{xy}}\, \mathbf{I}_m \\ \sqrt{\mathbf{n}_i^*} \end{pmatrix} \mathbf{U} \\ \mathbf{I} \end{bmatrix} \right\}^{-1} \mathbf{U}^T$$

The QR decomposition may be employed to perform the required matrix inversion.

A number of methods are available for estimating $\boldsymbol{\Sigma}_e$ and $\boldsymbol{\Sigma}_v$. If $n_i^x = n_i^y = 0$ for all i, then $n_i^{xy} = n_i$ and the method of moments estimators are:

$$\hat{\boldsymbol{\Sigma}}_e = \frac{1}{n-t} \sum_{i=1}^{t} \sum_{j=1}^{n_i} (\mathbf{u}_{ij} - \bar{\mathbf{u}})(\mathbf{u}_{ij} - \bar{\mathbf{u}})^T$$

and

$$\hat{\boldsymbol{\Sigma}}_v = \left[n - \sum_{i=1}^{t} n_i^2/n \right] \left[\sum_{i=1}^{t} n_i (\bar{\mathbf{u}}_i - \bar{\mathbf{u}})(\bar{\mathbf{u}}_i - \bar{\mathbf{u}})^T - (t-1)\hat{\boldsymbol{\Sigma}}_e \right].$$

Anderson et al. (1986) and Remadi and Amemiya (1994) discuss properties of maximum likelihood (ML) and restricted maximum likelihood (REML) estimators for multivariate components of variance for the balanced case (here, when $n_i^x = n_i^y = 0$ and $n_i^{xy} = n_j^{xy}$ for all i and j).

Properties of estimators for the covariance components in the general unbalanced case follow from the general mixed model discussed in Pinheiro and Bates (2000); these are the subject of a forthcoming paper. The package NLME (Pinheiro and Bates, 2000) for R and S-PLUS can be adapted to estimate $\mathbf{\Sigma}_e$ and $\mathbf{\Sigma}_v$ from the data using ML or REML, and to calculate the BLUPs for the small areas. S-PLUS code for small area estimation using NLME may be obtained from the authors.

5 Simulation results

To study the small sample efficiency of the estimators, we performed a simulation study with $m = 2$. A factorial design was employed, with factors

1. t: 10 or 20

2. Sample sizes (replicated if $t = 20$):

   (a) $\mathbf{n}^{xy} = (10, 10, 10, 10, 10, 10, 10, 10, 10, 10)$, $\mathbf{n}^x = \mathbf{0}$, $\mathbf{n}^y = \mathbf{0}$

   (b) $\mathbf{n}^{xy} = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$, $\mathbf{n}^x = (9, 8, 7, 6, 5, 4, 3, 2, 1, 0)$, $\mathbf{n}^y = \mathbf{0}$

   (c) $\mathbf{n}^{xy} = (0, 0, 0, 0, 5, 5, 5, 8, 8, 8)$, $\mathbf{n}^x = (10, 8, 6, 4, 5, 3, 1, 2, 1, 0)$, $\mathbf{n}^y = (0, 2, 4, 6, 0, 2, 4, 0, 1, 2)$

   (d) $\mathbf{n}^{xy} = (5, 4, 3, 2, 1, 0, 3, 2, 1, 0)$, $\mathbf{n}^x = (0, 1, 2, 3, 4, 5, 1, 2, 2, 3)$, $\mathbf{n}^y = (0, 0, 0, 0, 0, 0, 1, 1, 2, 2)$

3. $\mathbf{\Sigma}_v$ and $\mathbf{\Sigma}_e$: A=$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, B=$\begin{bmatrix} 1 & .3 \\ .3 & 1 \end{bmatrix}$, C=$\begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$, D=$\begin{bmatrix} 1 & -.5 \\ -.5 & 1 \end{bmatrix}$,
   E=$\begin{bmatrix} 1 & .6 \\ .6 & 4 \end{bmatrix}$, F=$\begin{bmatrix} 4 & .6 \\ .6 & 1 \end{bmatrix}$, G=$\begin{bmatrix} 1 & 1.8 \\ 1.8 & 4 \end{bmatrix}$, H=$\begin{bmatrix} 4 & 1.8 \\ 1.8 & 1 \end{bmatrix}$.

Each simulation was performed with 1000 iterations. Restricted maximum likelihood was used to estimate $\mathbf{\Sigma}_e$ and $\mathbf{\Sigma}_v$, using NLME (Pinheiro and Bates, 2000). All computations were performed in S-PLUS 2000 on a PC.

Selected results are given in Tables 1–6 for the four designs. Table 1 gives results for the balanced design (a). As expected, when $\mathbf{\Sigma}_v = \mathbf{\Sigma}_e$, the estimator from (7) has empirical MSE similar to that of the corresponding univariate estimator, but the estimator from (14) has greater empirical MSE because of the additional variability due to estimating all components of $\mathbf{\Sigma}_v$ and $\mathbf{\Sigma}_e$. As in Datta et al. (1999), the greatest gains in efficiency are achieved when the correlations ρ_v and ρ_e have opposite signs.

The biases of all estimators are negligible and hence not reported. In every simulation, the empirical MSE of the estimator (7) is within 0.01 of the theoretical value $\mathbf{D}_{iyy}$; $\mathbf{D}_{iyy}$ is consequently not reported in Tables 2–6.

Tables 2–6 give the simulation results for the three unbalanced designs. Not surprisingly, they show that the gains in efficiency are greater for each covariance structure when n_i^x is larger. Gains in efficiency are relatively modest when $n_i^x = 0$, except in the somewhat unrealistic case with $\mathbf{\Sigma}_{exy}$ and $\mathbf{\Sigma}_{vxy}$ have opposite signs. When n_i^x is large and ρ_v^2 is large, however, the estimator developed in this paper greatly improves the accuracy over the univariate estimator. This is true regardless of the values in $\mathbf{\Sigma}_e$.

6 Discussion

The results in this paper allow use of auxiliary information from either the same survey or a different survey to improve estimation in small areas. The estimator and its derived mean squared error account for the error in the auxiliary information.

For the case of sampling on two occasions with only partially overlapping information, such as occurs with a rotating panel survey, we would expect that ρ_e and ρ_v would both be large and positive. In this situation, we expect the greatest improvement over the univariate estimator when a large fraction of the sample is rotated out between the two time periods.

When using this estimator with multiple surveys, in most cases it will not be necessary to match sample observations between the two surveys. Even when the survey designs share the same primary sampling units, it is unlikely that the same persons are included in the surveys. Thus, it is overwhelmingly probable that in most small areas, $n_i^{xy} = 0$. Consequently, the estimator in (7) will involve $\mathbf{\Sigma}_v$ and $\mathbf{\Sigma}_e^*$ but not $\mathbf{\Sigma}_{exy}$. The vector $\mathbf{\Sigma}_{exy}$ is the only quantity, however, whose estimation requires that units in the two surveys be matched. The matrix $\mathbf{\Sigma}_e^*$ can be estimated from the two separate surveys, and $\mathbf{\Sigma}_v$ can be estimated provided that the number of small areas that contain observations from both surveys is sufficiently large.

As long as the nonresponse mechanism is noninformative, missing data are easily handled by the multivariate approach in this paper. Missing values of y may be treated as though the observation is not in $\mathcal{S}_{iy}$; missing values of $\mathbf{x}$ do not contribute to the prediction.

The results in this paper were presented for two surveys. They are easily extended to any number of surveys providing auxiliary information. They are also easily extended to the situation in which y is a multivariate response.

Appendix: Proofs

Several of the proofs rely on relations among the matrices defined in Section 2. These relations are given in the following lemmas for easy reference.

Lemma 1 *(Binomial Inverse Theorem) Let* $\mathbf{F}$ *and* $\mathbf{G}$ *be nonsingular matrices of dimension* $j \times j$ *and* $l \times l$*, respectively, and let the matrices* $\mathbf{H}$ *and* $\mathbf{L}$ *have dimensions* $j \times l$ *and* $l \times j$ *respectively. Then*

$$(\mathbf{F} + \mathbf{HGL})^{-1} = \mathbf{F}^{-1} - \mathbf{F}^{-1}\mathbf{H}(\mathbf{G}^{-1} + \mathbf{LF}^{-1}\mathbf{H})^{-1}\mathbf{LF}^{-1}.$$

Lemma 2 *For* $\mathbf{E}_i$ *and* $\mathbf{D}_i$ *defined in (5) and (6),*

$$\mathbf{D}_i^{-1} - \mathbf{E}_i = \boldsymbol{\Sigma}_v^{-1}. \tag{18}$$

$$\mathbf{E}_i\mathbf{D}_i = \mathbf{I}_m - \boldsymbol{\Sigma}_v^{-1}\mathbf{D}_i. \tag{19}$$

We also use the following lemma for Theorems 1 and 3.

Lemma 3 *For* $\mathbf{V}_i$, $\mathbf{R}_i$*, and* $\mathbf{Z}_i$ *defined in Section 2,*

$$\mathbf{V}_i^{-1} = \mathbf{R}_i^{-1} - \mathbf{R}_i^{-1}\mathbf{Z}_i\mathbf{D}_i\mathbf{Z}_i^T\mathbf{R}_i^{-1}. \tag{20}$$

In addition,

$$\mathbf{Z}_i^T\mathbf{R}_i^{-1} = \left[\mathbf{1}_{n_i^{xy}}^T \bigotimes \boldsymbol{\Sigma}_e^{-1}, \quad \mathbf{1}_{n_i^x}^T \bigotimes (\boldsymbol{\Sigma}_e^*)^{-1} \begin{pmatrix} \mathbf{I}_k \\ 0 \end{pmatrix}, \quad \mathbf{1}_{n_i^y}^T \bigotimes (\boldsymbol{\Sigma}_e^*)^{-1} \begin{pmatrix} \mathbf{0}_k \\ 1 \end{pmatrix}\right] \tag{21}$$

Let $\mathbf{m}_i^T = [\mathbf{m}_{i1},\ \mathbf{m}_{i2},\ \ldots\ \mathbf{m}_{it}]$ *be the* $m \times (tm)$ *matrix with* $\mathbf{m}_{ii} = \mathbf{I}_m$ *and all other entries 0. Then*

$$\mathbf{m}_i^T\mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1} = \boldsymbol{\delta}_i^T \bigotimes \mathbf{I}_m \bigotimes \boldsymbol{\Sigma}_v\mathbf{Z}_i^T\mathbf{V}_i^{-1} = \boldsymbol{\delta}_i^T \bigotimes \mathbf{I}_m \bigotimes \mathbf{D}_i\mathbf{Z}_i^T\mathbf{R}_i^{-1}. \tag{22}$$

Proof of Lemma 3: Equation (20) follows from Lemma 1. Equation (22) follows from direct calculation and (19), noting that

$$\mathbf{\Sigma}_v \mathbf{Z}_i^T \mathbf{V}_i^{-1} = \mathbf{\Sigma}_v \mathbf{Z}_i^T (\mathbf{R}_i^{-1} - \mathbf{R}_i^{-1}\mathbf{Z}_i \mathbf{D}_i \mathbf{Z}_i^T \mathbf{R}_i^{-1}) = \mathbf{\Sigma}_v (\mathbf{I}_m - \mathbf{E}_i \mathbf{D}_i)\mathbf{Z}_i^T \mathbf{R}_i^{-1} = \mathbf{D}_i \mathbf{Z}_i^T \mathbf{R}_i^{-1}.$$

Proof of Theorem 1: From Henderson (1975), the BLUP for $\boldsymbol{\mu}_i$ is

$$\tilde{\boldsymbol{\mu}}_i = \boldsymbol{\mu} + \mathbf{m}_i^T \mathbf{G}\mathbf{Z}^T \mathbf{V}^{-1}[\mathbf{u} - \mathbf{A}\boldsymbol{\mu}].$$

Equation (22) implies that

$$\mathbf{m}_i^T \mathbf{G}\mathbf{Z}^T \mathbf{V}^{-1}[\mathbf{u} - \mathbf{A}\boldsymbol{\mu}] = \mathbf{D}_i \mathbf{Z}_i^T \mathbf{R}_i^{-1}[\mathbf{u}_i - \mathbf{A}_i\boldsymbol{\mu}].$$

Result (7) is then proven by noting that

$$\mathbf{Z}_i^T \mathbf{R}_i^{-1}[\mathbf{u}_i - \mathbf{A}_i\boldsymbol{\mu}] = \sum_{j=1}^{n_i^{xy}} \mathbf{\Sigma}_e^{-1}[\mathbf{u}_{ij} - \boldsymbol{\mu}] + \sum_{j=n_i^{xy}+1}^{n_i^{xy}+n_i^x} \mathbf{\Sigma}_{exx}^{-1}[\mathbf{x}_{ij} - \boldsymbol{\mu}_x] + \sum_{j=n_i^{xy}+n_i^x+1}^{n_i^{xy}+n_i^x+n_i^y} \mathbf{\Sigma}_{eyy}^{-1}[y_{ij} - \mu_y].$$

To prove (8), note that

$$\begin{aligned}
\text{MSE}\,[\tilde{\boldsymbol{\mu}}_i] &= V[\tilde{\boldsymbol{\mu}}_i - \mathbf{v}_i] \\
&= V\left[(\mathbf{D}_i\mathbf{E}_i - \mathbf{I})\,\mathbf{v}_i + \mathbf{D}_i\left\{n_i^{xy}\mathbf{\Sigma}_e^{-1}\bar{\mathbf{e}}_{ixy} + \mathbf{n}_i^*(\mathbf{\Sigma}_e^*)^{-1}\begin{bmatrix}\bar{\mathbf{e}}_{ix}\\ \bar{\mathbf{e}}_{iy}\end{bmatrix}\right\}\right] \\
&= (\mathbf{D}_i\mathbf{E}_i - \mathbf{I})\mathbf{\Sigma}_v(\mathbf{E}_i\mathbf{D}_i - \mathbf{I}) + \mathbf{D}_i\left\{n_i^{xy}\mathbf{\Sigma}_e^{-1} + \mathbf{n}_i^*(\mathbf{\Sigma}_e^*)^{-1}\right\}\mathbf{D}_i \\
&= \mathbf{D}_i\mathbf{\Sigma}_v^{-1}\mathbf{D}_i + \mathbf{D}_i\mathbf{E}_i\mathbf{D}_i \\
&= \mathbf{D}_i.
\end{aligned}$$

The last equalities follow from (18) and (19).

Proof of Theorem 2:

Inversion of the partitioned matrices gives

$$\mathbf{\Sigma}_v^{-1} = \begin{bmatrix} \mathbf{\Sigma}_{vxx}^{-1} + (1-\rho_v^2)^{-1}\mathbf{\Sigma}_{vyy}\mathbf{c}\mathbf{c}^T & -(1-\rho_v^2)^{-1}\mathbf{c} \\ -(1-\rho_v^2)^{-1}\mathbf{c}^T & (1-\rho_v^2)^{-1}\mathbf{\Sigma}_{vyy}^{-1} \end{bmatrix}$$

and

$$\mathbf{E}_i = \begin{bmatrix} (n_i^{xy} + n_i^x)\mathbf{\Sigma}_{exx}^{-1} + n_i^{xy}(1-\rho_e^2)^{-1}\mathbf{\Sigma}_{eyy}\mathbf{d}\mathbf{d}^T & -n_i^{xy}(1-\rho_e^2)^{-1}\mathbf{d} \\ -n_i^{xy}(1-\rho_e^2)^{-1}\mathbf{d}^T & [n_i^y + n_i^{xy}(1-\rho_e^2)^{-1}]\mathbf{\Sigma}_{eyy}^{-1} \end{bmatrix}.$$

Consequently,

$$\mathbf{D}_i^{-1} = \mathbf{\Sigma}_v^{-1} + \mathbf{E}_i = \begin{bmatrix} \mathbf{M} & -\mathbf{h} \\ -\mathbf{h}^T & (1-\rho_v^2)^{-1}\mathbf{\Sigma}_{vyy}^{-1} + [n_i^y + n_i^{xy}(1-\rho_e^2)^{-1}]\mathbf{\Sigma}_{eyy}^{-1} \end{bmatrix}$$

and equation (9) follows by direct calculation.

Proof of Corollary 3:

By construction, $\mathbf{P}^T\mathbf{P} = \mathbf{\Sigma}_{vxx}$ and $\mathbf{P}^T\mathbf{\Lambda}\mathbf{P} = \mathbf{\Sigma}_{exx}$, so

$$\mathbf{M} = \mathbf{P}^{-1}(\mathbf{A} + \mathbf{\Lambda}^{-1/2}\mathbf{B}\mathbf{\Lambda}^{-1/2})\mathbf{P}^{-T},$$

where

$$\mathbf{A} = \mathbf{I} + (1-\rho_v^2)^{-1}\mathbf{P}\mathbf{c}\mathbf{\Sigma}_{vyy}\mathbf{c}^T\mathbf{P}^T$$

and

$$\mathbf{B} = (n_i^{xy} + n_i^x)\mathbf{I} + n_i^{xy}(1-\rho_e^2)^{-1}\mathbf{\Lambda}^{1/2}\mathbf{P}\mathbf{d}\mathbf{\Sigma}_{eyy}\mathbf{d}^T\mathbf{P}^T\mathbf{\Lambda}^{1/2}.$$

Since $\rho_v^2 = \mathbf{\Sigma}_{vyy}\mathbf{c}^T\mathbf{\Sigma}_{vxx}\mathbf{c}$ and $\rho_e^2 = \mathbf{\Sigma}_{eyy}\mathbf{d}^T\mathbf{\Sigma}_{exx}\mathbf{d}$, it follows that

$$\mathbf{A}^{-1} = \mathbf{I} - \mathbf{\Sigma}_{vyy}\mathbf{P}\mathbf{c}\mathbf{c}^T\mathbf{P}^T$$

and

$$\mathbf{B}^{-1} = (n_i^{xy} + n_i^x)^{-1}\left\{\mathbf{I} - n_i^{xy}\mathbf{\Sigma}_{eyy}[n_i^{xy} + n_i^x(1-\rho_e^2)]^{-1}\mathbf{\Lambda}^{1/2}\mathbf{P}\mathbf{d}\mathbf{\Sigma}_{eyy}\mathbf{d}^T\mathbf{P}^T\mathbf{\Lambda}^{1/2}\right\}.$$

Thus $\mathbf{c}^T\mathbf{P}^T\mathbf{A}^{-1} = (1-\rho_v^2)\mathbf{c}^T\mathbf{P}^T$ and $\mathbf{d}^T\mathbf{P}^T\mathbf{\Lambda}^{1/2}\mathbf{B}^{-1} = [n_i^{xy} + n_i^x(1-\rho_e^2)]^{-1}(1-\rho_e^2)\mathbf{d}^T\mathbf{P}^T\mathbf{\Lambda}^{1/2}$.

Consequently, using Lemma 1,

$$\begin{aligned}
\mathbf{h}^T\mathbf{M}^{-1}\mathbf{h} &= \mathbf{h}^T\mathbf{P}^T(\mathbf{A} + \mathbf{\Lambda}^{-1/2}\mathbf{B}\mathbf{\Lambda}^{-1/2})^{-1}\mathbf{P}\mathbf{h} \\
&= (1-\rho_v^2)^{-2}\mathbf{c}^T\mathbf{P}^T(\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}^{-1}\mathbf{A}^{-1})\mathbf{P}\mathbf{c} \\
&\quad + (n_i^{xy})^2(1-\rho_e^2)^{-2}\mathbf{d}^T\mathbf{P}^T\mathbf{\Lambda}^{1/2}(\mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{\Lambda}^{1/2}\mathbf{C}^{-1}\mathbf{\Lambda}^{1/2}\mathbf{B}^{-1})\mathbf{\Lambda}^{1/2}\mathbf{P}\mathbf{d} \\
&\quad + 2n_i^{xy}(1-\rho_v^2)^{-1}(1-\rho_e^2)^{-1}\mathbf{c}^T\mathbf{P}^T(\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}^{-1}\mathbf{A}^{-1})\mathbf{P}\mathbf{D} \\
&= (1-\rho_v^2)^{-1}\mathbf{\Sigma}_{vyy}^{-1}\rho_v^2 - \mathbf{c}^T\mathbf{P}^T\mathbf{C}^{-1}\mathbf{P}\mathbf{c} \\
&\quad + \left\{\mathbf{\Sigma}_{eyy}[n_i^{xy} + n_i^x(1-\rho_e^2)](1-\rho_e^2)\right\}^{-1}(n_i^{xy})^2\rho_e^2 \\
&\quad - [n_i^{xy} + n_i^x(1-\rho_e^2)]^{-2}(n_i^{xy})^2\mathbf{d}^T\mathbf{P}^T\mathbf{\Lambda}\mathbf{C}^{-1}\mathbf{\Lambda}\mathbf{P}\mathbf{d} \\
&\quad +2\,[n_i^{xy} + n_i^x(1-\rho_e^2)]^{-1}n_i^{xy}\mathbf{c}^T\mathbf{P}^T\mathbf{C}^{-1}\mathbf{\Lambda}\mathbf{P}\mathbf{d} \\
&= (1-\rho_v^2)^{-1}\mathbf{\Sigma}_{vyy}^{-1}\rho_v^2 + \left\{\mathbf{\Sigma}_{eyy}[n_i^{xy} + n_i^x(1-\rho_e^2)](1-\rho_e^2)\right\}^{-1}(n_i^{xy})^2\rho_e^2 - \mathbf{b}^T\mathbf{C}^{-1}\mathbf{b}.
\end{aligned}$$

After algebraic manipulation, the mean squared error of the multivariate estimator in (9) may be written as

$$\mathbf{D}_{iyy} = \mathrm{MSE}\,[\tilde{\mu}_{iy}^{univ}] \left\{ 1 - \mathbf{D}_{iyy} \left[\frac{\rho_v^2}{1-\rho_v^2} \boldsymbol{\Sigma}_{vyy}^{-1} + n_i^{xy} \frac{\rho_e^2}{1-\rho_e^2} \boldsymbol{\Sigma}_{eyy}^{-1} - \mathbf{h}^T \mathbf{M}^{-1} \mathbf{h} \right] \right\}.$$

Substituting in the above expression for $\mathbf{h}^T\mathbf{M}^{-1}\mathbf{h}$ and simplifying completes the proof of the corollary.

Proof of Theorem 3: Write

$$\hat{\boldsymbol{\mu}}_i = \mathbf{t}(\hat{\boldsymbol{\theta}}, \mathbf{u}) = [\mathbf{t}_1(\hat{\boldsymbol{\theta}}, \mathbf{u}), \ldots, \mathbf{t}_m(\hat{\boldsymbol{\theta}}, \mathbf{u})]^T$$

and suppose that $\boldsymbol{\theta}$ is a p-vector. As in Kackar and Harville (1984), Taylor's theorem yields

$$\mathbf{t}(\hat{\boldsymbol{\theta}}, \mathbf{u}) = \mathbf{t}(\boldsymbol{\theta}, \mathbf{u}) + \mathbf{d}^T(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p(t^{-1/2}),$$

where

$$\mathbf{d}(\boldsymbol{\theta}) = \left[\frac{\partial \mathbf{t}_1(\boldsymbol{\theta}, \mathbf{u})}{\partial \boldsymbol{\theta}}, \ldots, \frac{\partial \mathbf{t}_p(\boldsymbol{\theta}, \mathbf{u})}{\partial \boldsymbol{\theta}} \right].$$

Then by the arguments in Theorem A.1 in Prasad and Rao (1990),

$$E\left[\mathbf{d}^T(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \mathbf{d}(\boldsymbol{\theta}) \right]_{kl} = \mathrm{tr}\left[\nabla_{ik} \mathbf{V} \nabla_{il}^T \mathrm{Cov}\,(\hat{\boldsymbol{\theta}}) \right],$$

where

$$\nabla_{ik} = \mathrm{col}_{1\le j\le p}\; \boldsymbol{\delta}_k^T \frac{\partial [\mathbf{m}_i^T \mathbf{G} \mathbf{Z}^T \mathbf{V}^{-1}]}{\partial \boldsymbol{\theta}_j} = \mathrm{col}_{1\le j\le p}\; \boldsymbol{\delta}_k^T \frac{\partial [\boldsymbol{\delta}_i^T \otimes \mathbf{I}_m \otimes \mathbf{D}_i \mathbf{Z}_i^T \mathbf{R}_i^{-1}]}{\partial \boldsymbol{\theta}_j}$$

for $\mathbf{m}_i$ defined in Lemma 3. Now, using (21),

$$\frac{\partial \mathbf{D}_i \mathbf{Z}_i^T \mathbf{R}_i^{-1}}{\partial \boldsymbol{\theta}_q} = \left[\mathbf{1}_{n_i^{xy}}^T \bigotimes \frac{1}{n_i^{xy}} \frac{\partial \mathbf{D}_i \boldsymbol{\Sigma}_e^{-1}}{\partial \boldsymbol{\theta}_q},\;\; \mathbf{1}_{n_i^{x}}^T \bigotimes \frac{1}{n_i^{x}} \frac{\partial \mathbf{D}_i (\boldsymbol{\Sigma}_e^*)^{-1}}{\partial \boldsymbol{\theta}_q} \begin{pmatrix} \mathbf{I}_k \\ 0 \end{pmatrix},\;\; \mathbf{1}_{n_i^{y}}^T \bigotimes \frac{1}{n_i^{y}} \frac{\partial \mathbf{D}_i (\boldsymbol{\Sigma}_e^*)^{-1}}{\partial \boldsymbol{\theta}_q} \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} \right].$$

Using (2), and noting that

$$\left(\frac{\partial \mathbf{D}_i \mathbf{Z}_i^T \mathbf{R}_i^{-1}}{\partial \boldsymbol{\theta}_q} \right) \mathbf{Z}_i = n_i^{xy} \frac{\partial \mathbf{D}_i \boldsymbol{\Sigma}_e^{-1}}{\partial \boldsymbol{\theta}_q} + \frac{\partial \mathbf{D}_i \mathbf{n}_i^* (\boldsymbol{\Sigma}_e^*)^{-1}}{\partial \boldsymbol{\theta}_q} = \frac{\partial \mathbf{D}_i \mathbf{E}_i}{\partial \boldsymbol{\theta}_q},$$

we have that

$$\frac{\partial \mathbf{D}_i \mathbf{Z}_i^T \mathbf{R}_i^{-1}}{\partial \boldsymbol{\theta}_q} \mathbf{V}_i \left(\frac{\partial \mathbf{D}_i \mathbf{Z}_i^T \mathbf{R}_i^{-1}}{\partial \boldsymbol{\theta}_r} \right)^T = \frac{\partial \mathbf{D}_i \mathbf{Z}_i^T \mathbf{R}_i^{-1}}{\partial \boldsymbol{\theta}_q} (\mathbf{R}_i + \mathbf{Z}_i \boldsymbol{\Sigma}_v \mathbf{Z}_i^T) \left(\frac{\partial \mathbf{D}_i \mathbf{Z}_i^T \mathbf{R}_i^{-1}}{\partial \boldsymbol{\theta}_r} \right)^T$$

$$= \quad n_i^{xy} \frac{\partial \mathbf{D}_i \boldsymbol{\Sigma}_e^{-1}}{\partial \boldsymbol{\theta}_q} \boldsymbol{\Sigma}_e \left(\frac{\partial \mathbf{D}_i \boldsymbol{\Sigma}_e^{-1}}{\partial \boldsymbol{\theta}_r} \right)^T + n_i^x \frac{\partial \mathbf{D}_i (\boldsymbol{\Sigma}_e^*)^{-1}}{\partial \boldsymbol{\theta}_q} \begin{bmatrix} \boldsymbol{\Sigma}_{exx} & 0 \\ 0 & 0 \end{bmatrix} \left(\frac{\partial \mathbf{D}_i (\boldsymbol{\Sigma}_e^*)^{-1}}{\partial \boldsymbol{\theta}_r} \right)^T$$
$$+ \; n_i^y \frac{\partial \mathbf{D}_i (\boldsymbol{\Sigma}_e^*)^{-1}}{\partial \boldsymbol{\theta}_q} \begin{bmatrix} 0 & 0 \\ 0 & \boldsymbol{\Sigma}_{eyy} \end{bmatrix} \left(\frac{\partial \mathbf{D}_i (\boldsymbol{\Sigma}_e^*)^{-1}}{\partial \boldsymbol{\theta}_r} \right)^T + \frac{\partial \mathbf{D}_i \mathbf{E}_i}{\partial \boldsymbol{\theta}_q} \boldsymbol{\Sigma}_v \left(\frac{\partial \mathbf{D}_i \mathbf{E}_i}{\partial \boldsymbol{\theta}_r} \right)^T,$$

which is equivalent to the expression in (16).

The specific entries in $\mathbf{G}_{jl}$ are obtained using the following results:

$$\frac{\partial \mathbf{D}_i \boldsymbol{\Sigma}_e^{-1}}{\partial [\boldsymbol{\Sigma}_v]_{ab}} = \mathbf{D}_i \boldsymbol{\Sigma}_v^{-1} \boldsymbol{\Delta}_{ab} \boldsymbol{\Sigma}_v^{-1} \mathbf{D}_i \boldsymbol{\Sigma}_e^{-1}.$$

$$\frac{\partial \mathbf{D}_i (\boldsymbol{\Sigma}_e^*)^{-1}}{\partial [\boldsymbol{\Sigma}_v]_{ab}} = \mathbf{D}_i \boldsymbol{\Sigma}_v^{-1} \boldsymbol{\Delta}_{ab} \boldsymbol{\Sigma}_v^{-1} \mathbf{D}_i (\boldsymbol{\Sigma}_e^*)^{-1}.$$

$$\frac{\partial (\mathbf{D}_i \mathbf{E}_i)}{\partial [\boldsymbol{\Sigma}_v]_{ab}} = \mathbf{D}_i \boldsymbol{\Sigma}_v^{-1} \boldsymbol{\Delta}_{ab} \boldsymbol{\Sigma}_v^{-1} \mathbf{D}_i \mathbf{E}_i$$

$$\frac{\partial \mathbf{D}_i \boldsymbol{\Sigma}_e^{-1}}{\partial [\boldsymbol{\Sigma}_e]_{ab}} = \mathbf{D}_i \left[n_i^{xy} \boldsymbol{\Sigma}_e^{-1} \boldsymbol{\Delta}_{ab} \boldsymbol{\Sigma}_e^{-1} \mathbf{D}_i + \mathbf{n}_i^* (\boldsymbol{\Sigma}_e^*)^{-1} \boldsymbol{\Delta}_{ab} (\boldsymbol{\Sigma}_e^*)^{-1} \mathbf{D}_i \eta_{ab} - \boldsymbol{\Sigma}_e^{-1} \boldsymbol{\Delta}_{ab} \right] \boldsymbol{\Sigma}_e^{-1}.$$

$$\frac{\partial \mathbf{D}_i (\boldsymbol{\Sigma}_e^*)^{-1}}{\partial [\boldsymbol{\Sigma}_e]_{ab}} = \mathbf{D}_i \left[n_i^{xy} \boldsymbol{\Sigma}_e^{-1} \boldsymbol{\Delta}_{ab} \boldsymbol{\Sigma}_e^{-1} \mathbf{D}_i + \mathbf{n}_i^* (\boldsymbol{\Sigma}_e^*)^{-1} \boldsymbol{\Delta}_{ab} (\boldsymbol{\Sigma}_e^*)^{-1} \mathbf{D}_i \eta_{ab} - (\boldsymbol{\Sigma}_e^*)^{-1} \boldsymbol{\Delta}_{ab} \eta_{ab} \right] (\boldsymbol{\Sigma}_e^*)^{-1}$$

$$\frac{\partial (\mathbf{D}_i \mathbf{E}_i)}{\partial [\boldsymbol{\Sigma}_e]_{ab}} = \mathbf{D}_i \left[n_i^{xy} \boldsymbol{\Sigma}_e^{-1} \boldsymbol{\Delta}_{ab} \boldsymbol{\Sigma}_e^{-1} + \mathbf{n}_i^* (\boldsymbol{\Sigma}_e^*)^{-1} \boldsymbol{\Delta}_{ab} (\boldsymbol{\Sigma}_e^*)^{-1} \eta_{ab} \right] (\mathbf{D}_i \mathbf{E}_i - \mathbf{I}).$$

References

Anderson, B.M., Anderson, T.W. and Olkin, I. (1986). Maximum likelihood estimators and likelihood criteria in multivariate components of variance. *Annals of Statistics*, 14, 405–417.

Battese, G.E., Harter, R.M., 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.

Cochran, W.G. (1977). *Sampling Techniques, 3rd ed.* New York: Wiley.

Datta, G.S., Day, B. and Basawa, I. (1999). Empirical best linear unbiased and empirical Bayes prediction in multivariate small area estimation. *Journal of Statistical Planning and Inference,* 75, 269–279.

Datta, G.S. and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, 10, 613-627.

Dempster, A.P., Rubin, D.B., and Tsutakawa, R.K. (1981). Estimation in covariance component models. *Journal of the American Statistical Association*, 76, 341-353.

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.

Fuller, W.A. and Harter, R.M. (1987). The multivariate components of variance model for small area estimation. In *Small Area Statistics*, ed. R. Platek, J.N.K. Rao, C.E. Särndal and M.P. Singh. New York: Wiley, 103–123.

Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science,* 9, 55–93.

Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics,* 31, 423–447.

Kackar, R.N. and Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association*, 79, 853–862.

Marker, D.A. (1999). Organization of small area estimators using a generalized linear regression framework. *Journal of Official Statistics,* 15, 1–24.

National Research Council (1999). *Small-area estimates of school-age children in poverty: Interim report 3.* C. F. Citro and G. Kalton, eds. Panel on Estimates of Poverty for Small Geographic Areas, Committee on National Statistics. Washington, D.C.: National Academy Press.

Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-effects models in S and S-PLUS.* New York: Springer-Verlag.

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.

Prasad, N.G.N. and Rao, J.N.K. (1999). On robust small area estimation using a simple random effects model. *Survey Methodology*, 25, 67–72.

Rao, J.N.K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, 25, 175–186.

Remadi, S. and Amemiya, Y. (1994). Asymptotic properties of the estimators for multivariate components of variance. *Journal of Multivariate Analysis*, 49, 110–131.

Wiersema, B., McDowall, D. and Loftin, C. (2000). Comparing metropolitan area estimates of crime from the National Crime Victimization Survey and Uniform Crime Reports. Paper presented at the Joint Statistical Meetings, Indianapolis, August 2000.

Table 1: Simulation results for the balanced design with $n_i^{xy} = 10$ for each small area. Each entry of the table is the Monte Carlo MSE of the small area mean.

t	$\mathbf{\Sigma}_v$	$\mathbf{\Sigma}_e$	MK	ME	UK	UE	TK
20	B	B	0.0901	0.0922	0.0901	0.0915	0.0909
20	C	B	0.0817	0.0858	0.0918	0.0935	0.0814
20	C	D	0.0543	0.0631	0.0903	0.0916	0.0544
20	C	A	0.0717	0.0769	0.0901	0.0914	0.0725
20	D	A	0.0891	0.0916	0.0917	0.0930	0.0885
20	E	A	0.0980	0.0985	0.0981	0.0985	0.0973
20	G	A	0.0926	0.0949	0.0985	0.0989	0.0913
20	F	A	0.0904	0.0932	0.0909	0.0928	0.0901
20	H	A	0.0671	0.0746	0.0903	0.0919	0.0677
20	H	B	0.0725	0.0789	0.0896	0.0914	0.0733

NOTE: MK = multivariate estimates, with $\mathbf{\Sigma}_v$ and $\mathbf{\Sigma}_e$ known; ME = multivariate estimates, with $\mathbf{\Sigma}_v$ and $\mathbf{\Sigma}_e$ estimated from the data; UK = univariate estimates, with variance components known; UE = univariate estimates, with variance components estimated from the data; TK = theoretical multivariate variance with $\mathbf{\Sigma}_v$ and $\mathbf{\Sigma}_e$ known (TK = $\mathbf{D}_{iyy}$).

Table 2: Simulation results for Design (b). MK, ME, UK, and UE are as in Table 1.

	n_i^{xy}	1	2	3	4	5	6	7	8	9	10
	n_i^{x}	9	8	7	6	5	4	3	2	1	0
	n_i^{y}	0	0	0	0	0	0	0	0	0	0
$t = 10$	MK	.21	.17	.17	.13	.12	.12	.11	.10	.09	.08
$\mathbf{\Sigma}_v$: C	ME	.29	.23	.22	.17	.15	.14	.13	.12	.10	.09
$\mathbf{\Sigma}_e$: B	UK	.48	.32	.26	.19	.17	.15	.13	.11	.10	.09
	UE	.56	.36	.29	.21	.18	.16	.14	.12	.11	.09
$t = 20$	MK	.21	.17	.16	.14	.12	.11	.10	.10	.09	.08
$\mathbf{\Sigma}_v$: C	ME	.24	.20	.18	.16	.13	.13	.11	.11	.09	.09
$\mathbf{\Sigma}_e$: B	UK	.47	.33	.24	.21	.16	.14	.12	.11	.10	.09
	UE	.51	.35	.25	.22	.17	.15	.13	.11	.10	.09
$t = 20$	MK	.17	.14	.11	.11	.11	.10	.10	.10	.09	.09
$\mathbf{\Sigma}_v$: C	ME	.18	.15	.12	.11	.11	.11	.10	.10	.09	.10
$\mathbf{\Sigma}_e$: C	UK	.49	.34	.24	.20	.17	.15	.13	.11	.10	.09
	UE	.53	.36	.25	.21	.18	.15	.13	.12	.10	.10
$t = 20$	MK	.18	.14	.12	.10	.08	.08	.07	.06	.06	.05
$\mathbf{\Sigma}_v$: C	ME	.23	.18	.14	.12	.10	.09	.08	.08	.07	.06
$\mathbf{\Sigma}_e$: D	UK	.47	.32	.25	.20	.17	.15	.12	.11	.10	.09
	UE	.51	.35	.27	.21	.18	.16	.12	.11	.10	.10
$t = 10$	MK	.18	.14	.11	.10	.08	.08	.07	.07	.06	.06
$\mathbf{\Sigma}_v$: C	ME	.28	.21	.15	.13	.11	.10	.09	.09	.08	.07
$\mathbf{\Sigma}_e$: D	UK	.51	.32	.25	.20	.16	.15	.13	.12	.10	.10
	UE	.62	.38	.26	.23	.18	.16	.13	.12	.10	.10
$t = 20$	MK	.46	.32	.23	.19	.17	.14	.12	.11	.10	.09
$\mathbf{\Sigma}_v$: B	ME	.51	.35	.25	.20	.18	.15	.12	.12	.11	.09
$\mathbf{\Sigma}_e$: B	UK	.48	.34	.24	.20	.17	.15	.12	.11	.10	.09
	UE	.53	.36	.26	.21	.18	.15	.12	.12	.10	.09

Table 3: Simulation results for Design (b), continued. MK, ME, UK, and UE are as in Table 1.

	n_i^{xy}	1	2	3	4	5	6	7	8	9	10
	n_i^{x}	9	8	7	6	5	4	3	2	1	0
	n_i^{y}	0	0	0	0	0	0	0	0	0	0
$t = 10$	MK	.53	.34	.24	.20	.16	.14	.13	.12	.10	.09
$\mathbf{\Sigma}_v$: G	ME	.64	.41	.28	.22	.18	.15	.14	.13	.11	.09
$\mathbf{\Sigma}_e$: A	UK	.77	.43	.29	.22	.19	.16	.14	.13	.11	.09
	UE	.81	.45	.31	.23	.18	.16	.15	.13	.11	.09
$t = 10$	MK	.77	.42	.32	.23	.19	.17	.15	.13	.11	.10
$\mathbf{\Sigma}_v$: E	ME	.87	.46	.35	.24	.20	.18	.16	.14	.11	.11
$\mathbf{\Sigma}_e$: A	UK	.80	.42	.32	.23	.19	.17	.15	.13	.11	.10
	UE	.85	.44	.33	.24	.20	.17	.15	.13	.11	.11
$t = 10$	MK	.49	.33	.26	.20	.17	.14	.12	.12	.10	.09
$\mathbf{\Sigma}_v$: F	ME	.62	.44	.31	.24	.19	.16	.13	.13	.11	.10
$\mathbf{\Sigma}_e$: A	UK	.52	.34	.27	.20	.17	.14	.12	.12	.10	.09
	UE	.61	.40	.29	.23	.18	.15	.13	.13	.11	.10
$t = 20$	MK	.21	.14	.12	.10	.10	.09	.09	.08	.09	.08
$\mathbf{\Sigma}_v$: A	ME	.23	.15	.12	.11	.11	.10	.09	.09	.09	.09
$\mathbf{\Sigma}_e$: C	UK	.48	.34	.24	.20	.17	.14	.12	.11	.10	.09
	UE	.51	.36	.25	.21	.17	.14	.12	.11	.10	.09

Table 4: Simulation results for Design (c). MK, ME, UK, and UE are as in Table 1.

	n_i^{xy}	0	0	0	0	5	5	5	8	8	8
	n_i^{x}	10	8	6	4	5	3	1	2	1	0
	n_i^{y}	0	2	4	6	0	2	4	0	1	2
$t = 10$	MK	.24	.17	.14	.12	.12	.10	.09	.09	.09	.08
$\boldsymbol{\Sigma}_v$: C	ME	.38	.24	.18	.14	.15	.12	.10	.11	.10	.09
$\boldsymbol{\Sigma}_e$: B	UK	.91	.33	.20	.15	.17	.12	.10	.12	.10	.09
	UE	1.07	.38	.23	.16	.18	.13	.11	.12	.11	.10
$t = 10$	MK	.89	.34	.21	.14	.16	.11	.10	.10	.09	.09
$\boldsymbol{\Sigma}_v$: B	ME	1.21	.41	.24	.15	.18	.13	.11	.11	.10	.10
$\boldsymbol{\Sigma}_e$: B	UK	.97	.34	.21	.14	.17	.11	.10	.10	.09	.09
	UE	1.12	.39	.23	.15	.18	.12	.11	.11	.09	.10
$t = 10$	MK	.25	.17	.14	.12	.08	.07	.06	.06	.06	.05
$\boldsymbol{\Sigma}_v$: C	ME	.41	.24	.17	.14	.11	.10	.08	.08	.07	.07
$\boldsymbol{\Sigma}_e$: D	UK	.90	.31	.20	.15	.16	.12	.09	.11	.10	.09
	UE	1.02	.36	.22	.16	.16	.13	.10	.12	.10	.09
$t = 10$	MK	3.81	.44	.24	.16	.18	.14	.11	.12	.11	.10
$\boldsymbol{\Sigma}_v$: E	ME	4.89	.45	.25	.16	.19	.15	.12	.12	.11	.11
$\boldsymbol{\Sigma}_e$: B	UK	4.10	.44	.24	.16	.18	.14	.12	.12	.11	.10
	UE	4.57	.45	.25	.16	.19	.14	.12	.12	.11	.10
$t = 10$	MK	1.03	.36	.20	.15	.17	.12	.10	.11	.10	.09
$\boldsymbol{\Sigma}_v$: G	ME	1.46	.44	.22	.16	.19	.13	.11	.12	.11	.10
$\boldsymbol{\Sigma}_e$: A	UK	3.82	.47	.23	.17	.19	.13	.11	.12	.11	.10
	UE	4.31	.50	.23	.17	.19	.13	.11	.12	.11	.10
$t = 10$	MK	.20	.15	.12	.10	.10	.09	.08	.07	.08	.07
$\boldsymbol{\Sigma}_v$: H	ME	.29	.22	.17	.13	.14	.11	.10	.10	.10	.09
$\boldsymbol{\Sigma}_e$: A	UK	.91	.34	.21	.15	.16	.12	.10	.10	.11	.09
	UE	1.04	.38	.23	.17	.18	.13	.11	.11	.11	.10

Table 5: Simulation results for Design (d). MK, ME, UK, and UE are as in Table 1.

	n_i^{xy}	5	4	3	2	1	0	3	2	1	0
	n_i^{x}	0	1	2	3	4	5	1	2	2	3
	n_i^{y}	0	0	0	0	0	0	1	1	2	2
$t = 10$	MK	.14	.17	.19	.22	.25	.32	.19	.19	.18	.21
$\boldsymbol{\Sigma}_v$: C	ME	.17	.21	.25	.32	.41	.50	.22	.24	.23	.27
$\boldsymbol{\Sigma}_e$: B	UK	.16	.21	.26	.34	.49	1.00	.23	.26	.25	.33
	UE	.18	.24	.31	.40	.60	1.20	.25	.28	.30	.42
$t = 20$	MK	.09	.09	.12	.15	.21	.32	.10	.13	.16	.22
$\boldsymbol{\Sigma}_v$: C	ME	.11	.12	.16	.20	.28	.43	.13	.16	.19	.26
$\boldsymbol{\Sigma}_e$: D	UK	.17	.20	.25	.36	.51	1.02	.19	.24	.26	.33
	UE	.18	.22	.27	.40	.56	1.12	.21	.26	.28	.37
$t = 20$	MK	.14	.16	.18	.21	.26	.32	.16	.19	.18	.22
$\boldsymbol{\Sigma}_v$: C	ME	.16	.19	.22	.26	.31	.42	.18	.22	.21	.26
$\boldsymbol{\Sigma}_e$: B	UK	.16	.20	.26	.34	.49	1.06	.20	.25	.24	.33
	UE	.18	.21	.27	.37	.54	1.14	.21	.28	.26	.38
$t = 20$	MK	.17	.17	.18	.20	.24	.32	.16	.18	.19	.22
$\boldsymbol{\Sigma}_v$: C	ME	.18	.18	.20	.22	.27	.37	.17	.19	.21	.24
$\boldsymbol{\Sigma}_e$: C	UK	.17	.20	.25	.33	.53	.99	.18	.25	.26	.32
	UE	.18	.21	.27	.38	.59	1.07	.20	.27	.28	.36

Table 6: Simulation results for Design (d), continued. MK, ME, UK, and UE are as in Table 1.

	n_i^{xy}	5	4	3	2	1	0	3	2	1	0
	n_i^{x}	0	1	2	3	4	5	1	2	2	3
	n_i^{y}	0	0	0	0	0	0	1	1	2	2
$t = 20$	MK	.13	.15	.20	.26	.41	1.29	.16	.21	.23	.39
$\mathbf{\Sigma}_v$: G	ME	.15	.17	.23	.30	.50	1.54	.18	.24	.24	.42
$\mathbf{\Sigma}_e$: D	UK	.18	.22	.32	.44	.78	3.98	.24	.31	.31	.46
	UE	.19	.23	.32	.45	.81	4.26	.24	.32	.31	.47
$t = 20$	MK	.19	.23	.27	.38	.60	1.33	.22	.27	.29	.39
$\mathbf{\Sigma}_v$: G	ME	.20	.24	.29	.41	.68	1.54	.23	.29	.30	.41
$\mathbf{\Sigma}_e$: B	UK	.19	.24	.30	.45	.80	4.10	.23	.29	.31	.45
	UE	.19	.25	.30	.46	.84	4.36	.23	.30	.32	.45
$t = 20$	MK	.08	.09	.11	.13	.16	.23	.10	.12	.13	.16
$\mathbf{\Sigma}_v$: H	ME	.12	.12	.15	.19	.22	.31	.14	.16	.17	.23
$\mathbf{\Sigma}_e$: D	UK	.17	.19	.25	.35	.51	.99	.20	.24	.25	.34
	UE	.18	.21	.27	.39	.56	1.07	.22	.26	.27	.38
$t = 20$	MK	.12	.13	.14	.17	.20	.23	.13	.14	.15	.16
$\mathbf{\Sigma}_v$: H	ME	.16	.16	.18	.22	.26	.30	.17	.18	.19	.20
$\mathbf{\Sigma}_e$: B	UK	.18	.20	.24	.35	.54	.98	.20	.25	.25	.33
	UE	.19	.21	.26	.38	.62	1.05	.21	.27	.28	.37
$t = 20$	MK	.17	.21	.26	.33	.48	.93	.20	.24	.24	.32
$\mathbf{\Sigma}_v$: B	ME	.18	.23	.28	.38	.58	1.09	.22	.27	.27	.36
$\mathbf{\Sigma}_e$: B	UK	.17	.21	.26	.36	.50	.99	.20	.25	.24	.33
	UE	.18	.22	.28	.39	.56	1.07	.21	.26	.27	.36