Stat Papers (2010) 51:497–509
DOI 10.1007/s00362-008-0136-1

REGULAR ARTICLE

Synthetic and composite estimation under a superpopulation model

Wojciech Niemiro · Jacek Wesołowski

Received: 12 October 2006 / Accepted: 17 March 2008 / Published online: 16 April 2008

Abstract Under a simple superpopulation model for an arbitrary sampling design we derive optimal linear unbiased estimators/predictors of a mean in a domain. They can be viewed as synthetic and composite estimators of small area estimation theory when no auxiliary variable is available. Moreover, we show that the only requirement for optimality of a sampling strategy is to use any sampling plan of fixed sample size together with traditional estimators (as designed for simple random sampling without replacement). Finally, for symmetric sampling plans, simplified formulas (based on the first two moments of sample sizes) for optimal synthetic and composite estimators and their MSE's are derived. Throughout the paper we consistently use the model-design setup.

Keywords Small area estimation · Model-design setup · Optimality of BLUE and BLUP

1 Introduction

The basic problem of small area estimation is concerned with providing reliable estimates for parameters of domains for which the sample size does not permit to use standard estimates provided by the survey sampling theory due to their prohibitive levels of errors.

Work partially supported by Polish Ministry of Science and Higher Education Grant No. 1 H02B 029 30.

W. Niemiro (✉)
Uniwersytet M. Kopernika, Toruń and Główny Urząd Statystyczny, Warszawa, Poland
e-mail: wniemiro@gmail.com

J. Wesołowski
Politechnika Warszawska and Główny Urząd Statystyczny, Warszawa, Poland
e-mail: wesolo@mini.pw.edu.pl

The simplest heuristic idea then is to believe that the parameters of small areas are the same as the respective parameters in the whole population and to use the estimators designed for the population parameters (or their simple transformations) to estimate respective small area quantities. This idea is an informal basis of so called synthetic estimators which goes back to Gonzalez (1973), see also Rao (2003, p. 46). In the case the population parameters and small area parameters are different such an approach may lead to heavily biased estimators, therefore the heuristic approach is to take a weighted average of the estimate for the whole population and the design based estimate for the small area. Such a procedure is often referred to as "borrowing the strength" and leads to so called composite estimators. For more information on synthetic and composite estimation we refer to Ghosh (2001), Rao (2003, Sects. 4.2 and 4.3) or a review paper by Marker (1999, Sects. 2.1.6 and 2.1.7).

In this note our aim is to present rigorous derivations of synthetic and composite estimators under a superpopulation model combined with arbitrary non-informative sampling designs. The model we consider here is described in Sect. 2. It is very simple, with no auxiliary variables and with no autocorrelation, as considered e.g. in the classical monograph Cassel et al. (1977, Sect. 5.3) or in lecture notes of Ghosh (2001, Sect. 0.3). The very simplicity of the model makes it a good starting point for a discussion of relations between model- and design-based approaches. In our paper, we consistently use the model-design setup: in defining the class of linear and unbiased estimators as well as in the criterion of optimality (MSE).

Many authors considered similar problems of optimal unbiased linear estimation/prediction under purely model-based approaches. Standard references in this setting are several papers of Royall and co-authors starting with Royall (1970). A review can be found in Cassel et al. (1977, Chap. 5). For more recent results one can consult e.g. Bellhouse (1984) or Chambers (2005).

The task of deriving optimal strategy is split into two stages.

First, under an arbitrary sampling design we find optimal linear unbiased (synthetic) estimators and predictors (composite estimators). This is the main result of Sect. 3. The derivation is based on a technical lemma on constrained minimization given in Appendix A.1. In the special case of fixed size sampling designs our formulas agree with those known in the literature, e.g. in Scott and Smith (1969), Cassel et al. (1977, Chap. 5) or Ghosh (2001, Sect. 0.3).

Second, we find the family of optimal sampling designs. The main result of Sect. 4 says that the only requirement for optimality of a sampling strategy is to use BLUE and BLUP together with any design of a fixed sample size. The argument is based on inequalities for functionals of positive definite matrices which are proved in Appendix A.2.

Let us emphasize that the problem we solve is of the same nature as e.g. those considered in Royall (1970) or Cassel et al. (1977, Chaps. 4 and 5). However, we consider optimality within a different class of estimators. Among the distinctive features of our approach the most notable is the definition of linearity (we consider estimators/predictors with weights which do not depend on the sample).

Despite the fact that designs with random domain sample size are not optimal, they often occur in practice. The general formulas for BLUE and BLUP and their MSE's for any sampling design are analytically simple but they can be difficult to apply numerically, due to the fact that they involve inverting large matrices. However, when the attention is restricted to symmetric designs, the respective formulas are quite elementary functions of the first two moments of the sample sizes. They are derived in Sect. 5.

2 The model-design setup

Denote a population by $U = \{1, \ldots, N\}$. Consider a sampling plan (p) on U with (π_k) and (π_{kl}) being inclusion probabilities, respectively, of the first and second order. Denote $\pi = (\pi_1, \ldots, \pi_N)^T$ and $\Pi = \text{diag}(\pi)$, i.e. $\Pi\mathbf{1} = \pi$, where $\mathbf{1} = (1, \ldots, 1)^T$. Also let $P = [\pi_{ij}]$ (in particular $\pi_{ii} = \pi_i$). Let S denote the sample drawn from U according to (p). Let $I_k(S) = 1$ if $k \in S$ and $I_k(S) = 0$ if $k \in U \setminus S$. Let $\mathbf{I} = (I_1(S), \ldots, I_N(S))^T$. Then $\mathbb{E}(\mathbf{I}) = \pi$ and the covariance matrix of $\mathbf{I}$ is $P - \pi\pi^T$.

We consider the following superpopulation model (Ghosh 2001, p. 57). The values of the variable of interest corresponding to all the population units are regarded as realizations of random variables $Y_1, \ldots, Y_N$. From now on we assume that $\mathbf{Y} = (Y_1, \ldots, Y_N)^T$ is a random vector with square integrable independent identically distributed components with the expectation $\mu > 0$ and the coefficient of variation γ. Recall that $\gamma = \sigma/\mu$, where σ^2 is the common variance of Y_is. Thus $\mathbb{E}(\mathbf{Y}) = \mu\mathbf{1}$ and the covariance matrix of $\mathbf{Y}$ is $\sigma^2 I$, where I denotes the identity matrix.

We assume that the random vectors $\mathbf{Y}$ and $\mathbf{I}$ are independent (the sampling plan is noninformative). Consider the random vector $\mathbf{Z} = \text{diag}(\mathbf{I})\mathbf{Y}$, i.e. $\mathbf{Z} = (Z_1, \ldots, Z_N)^T = (Y_1 I_1(S), \ldots, Y_N I_N(S))^T$. Thus we regard $\mathbf{Z}$ as the vector of observations while the random vector $\mathbf{Y}$ is not fully observed. In the calculations below we are using basic properties of conditioning. Note that

$$\mathbb{E}(\mathbf{Z}) = \mathbb{E}\left(\text{diag}(\mathbf{I})\mathbb{E}(\mathbf{Y}|\mathbf{I})\right) = \mu\pi.$$

Let us stress that the symbol $\mathbb{E}$ in the above formula (and indeed in all the formulas in this paper) denotes expectation with respect both to the sampling plan (p) and the superpopulation model $\mathbf{Y}$. Thus for a function f,

$$\mathbb{E}f(\mathbf{Y}, S) = \int_{\mathbb{R}^N} \sum_{s \subset U} f(\mathbf{y}, s)\ p(s)\ \mathbb{P}_{\mathbf{Y}}(d\mathbf{y}),$$

where $\mathbb{P}_{\mathbf{Y}}$ denotes the probability distribution of the random vector $\mathbf{Y}$. Note that in some papers, frequently cited in the literature, instead of $\mathbb{E}$ authors use a notation of the type $E_\xi E_p$, where E_ξ stands for the expectation with respect to the model while E_p stands for the expectation with respect to the design.

Let V denote the covariance matrix of the random vector $\mathbf{Z}$, i.e.

$$\begin{aligned} V &= \mathbb{E}\left[(\mathbf{Z} - \mu\pi)(\mathbf{Z} - \mu\pi)^T\right] \\ &= \mathbb{E}\left\{[diag(\mathbf{I})(\mathbf{Y} - \mu\mathbf{1}) + \mu(\mathbf{I} - \pi)]\,[diag(\mathbf{I})(\mathbf{Y} - \mu\mathbf{1}) + \mu(\mathbf{I} - \pi)]^T\right\} \\ &= \mathbb{E}\left[diag(\mathbf{I})\mathbb{E}\left[(\mathbf{Y} - \mu\mathbf{1})(\mathbf{Y} - \mu\mathbf{1})^T|\mathbf{I}\right]diag(\mathbf{I})\right] + \mu^2\mathbb{E}\left[(\mathbf{I} - \pi)(\mathbf{I} - \pi)^T\right]. \end{aligned}$$

Consequently

$$V = \sigma^2 \Pi + \mu^2 (P - \pi\pi^T). \tag{1}$$

The parameters of our model are μ and γ. We assume that γ is known and μ is unknown. Therefore, we prefer to write formulas in terms of the matrix

$$\tilde{V} = \frac{V}{\sigma^2} = \Pi + \gamma^{-2}(P - \pi\pi^T). \tag{2}$$

For future reference, let us note that V and thus also $\tilde{V}$ are invertible. It follows from the fact that Π is positive definite (under the standard assumption that all π_is are strictly positive) and $P - \pi\pi^T$ is semi-positive definite.

Consider an arbitrary domain $D \subset U$. We are interested in the domain mean, $\bar{Y}_D = \frac{1}{N_D} \sum_{i \in D} Y_i$, where $N_D = \#(D)$ denotes the cardinality of D. It will be convenient to write $\bar{Y}_D = \frac{1}{N_D} \mathbf{Y}^T \mathbf{e}_D$, where $\mathbf{e}_D = (e_1, \ldots, e_N)$ with $e_i = 1$ if $i \in D$ and $e_i = 0$ if $i \notin D$. Notice that in our setting $\bar{Y}_D$ is a random variable so we should rather speak of predicting, not estimating this quantity. The auxiliary task, which is perhaps also of independent interest, is to estimate the superpopulation (model) mean μ.

In Sect. 3, for a fixed but arbitrary sampling plan (p), we will consider the problem of deriving the best linear unbiased estimator (BLUE) $\hat{\mu}$ of μ and the best linear unbiased predictor (BLUP) $\widehat{Y}_D$ of $\bar{Y}_D$. Let us stress that by linear estimators/predictors we mean statistics which are linear functions of the observed variables Z_i, i.e. they are of the form

$$\mathbf{w}^T \mathbf{Z} = \sum_{i \in U} w_i Z_i,$$

where the coefficients w_i are fixed numbers (they do not depend on the sample S). Note that we can write $\mathbf{w}^T \mathbf{Z} = \sum_{i \in S} w_i Y_i$. Our approach and terminology used here are different from those prevailing in the model-based literature: for example Scott and Smith (1969) as well as Royall (1970) consider, under the name of "linear estimators", a different family of statistics (namely those of the form $\sum_{i \in S} w_i(S) Y_i$, where the weights are allowed to depend on S). Note also that the purely design-based linearity may be understood in yet another sense (namely as linearity with respect to the indicator variables I_i) and in such a setting the π-estimator is the only linear unbiased one.

3 Optimal estimation and prediction

As it has been already announced, in this section we derive optimal linear unbiased estimators/predictors for an arbitrary fixed sampling plan. At the same time, it is the first step in deriving optimal strategy, which consists of a pair: an estimator/predictor and a sampling plan. These estimation and prediction problems can be embedded in a minimization scheme which is solved in Lemma 1 (see Appendix A.1). Therefore Lemma 1 is the basic tool of the proof of Theorem 1 given below.

Theorem 1 *Let* (p) *be an arbitrary sampling plan.*
The best linear unbiased estimator (BLUE) $\hat{\mu}$ *of* μ *has the form*

$$\hat{\mu} = \frac{\pi^T \tilde{V}^{-1} \mathbf{Z}}{\pi^T \tilde{V}^{-1} \pi}. \tag{3}$$

The variance of BLUE is

$$Var(\hat{\mu}) = \frac{\sigma^2}{\pi^T \tilde{V}^{-1} \pi}. \tag{4}$$

The best linear unbiased predictor (BLUP) $\widehat{Y}_D$ *of* $\bar{Y}_D$ *has the form*

$$\widehat{Y}_D = \frac{1}{N_D} \left[\frac{N_D - \pi_D^T \tilde{V}^{-1} \pi}{\pi^T \tilde{V}^{-1} \pi} \pi^T \tilde{V}^{-1} \mathbf{Z} + \pi_D^T \tilde{V}^{-1} \mathbf{Z} \right], \tag{5}$$

where $\pi_D = \Pi \mathbf{e}_D$ *and* $N_D = \#(D)$*. The MSE of BLUP is*

$$\mathbb{E} \left(\widehat{Y}_D - \bar{Y}_D \right)^2 = \frac{\sigma^2}{N_D^2} \left[\frac{(N_D - \pi_D^T \tilde{V}^{-1} \pi)^2}{\pi^T \tilde{V}^{-1} \pi} - \pi_D^T \tilde{V}^{-1} \pi_D + N_D \right]. \tag{6}$$

Proof In the case of BLUE we take $X = \mu$ in Lemma 1. Then $\mathbf{r} = \pi$, $\mathbf{c} = 0$, $\Sigma = \sigma^2 \tilde{V}$ and the formulas (3) and (4) follow from (22) and (21), respectively.

In the case of BLUP we take $X = \bar{Y}_D$ in Lemma 1. Then $\mathbf{r} = \pi$, $\mathbf{c} = \frac{\sigma^2}{N_D} \pi_D$, $\Sigma = \sigma^2 \tilde{V}$ and the formulas (6) and (5) follow from (22) and (21), respectively. □

Remark 1 Let us note that for $D = U$ the formulas (5) and (6) reduce to

$$\widehat{Y}_U = \frac{\pi^T \tilde{V}^{-1} \mathbf{Z}}{\pi^T \tilde{V}^{-1} \pi} \quad \text{and} \quad \mathbb{E} \left(\widehat{Y}_U - \bar{Y}_U \right)^2 = \sigma^2 \left(\frac{1}{\pi^T \tilde{V}^{-1} \pi} - \frac{1}{N} \right). \tag{7}$$

To the best of our knowledge even in the simplified situation when $D = U$, the formulas in (7), which hold for any sampling design, are new. In this context, apart from the results which cover only the case of designs with fixed sample sizes, we should mention Theorem 5.4 of Cassel et al. (1977), which asserts that $\mathbb{E}(T - \bar{Y}_U)^2 \geq \sigma^2 \mathbb{E} \left(\frac{1}{n(S)} - \frac{1}{N} \right)$ for any model-unbiased (not necessarily linear in any sense) predictor T.

In the remaining part of this section we specialize the general result of Theorem 1 to the case of fixed sample sizes.

Corollary 1 *Assume that* (p) *is a sampling design with the fixed (nonrandom) sample size* $\#S = n$*.*

(i) *Then the BLUE assumes the form*

$$\hat{\mu} = \frac{1}{n} \sum_{i \in S} Y_i \tag{8}$$

and its variance equals

$$Var(\hat{\mu}) = \frac{\sigma^2}{n}. \tag{9}$$

(ii) *Assume additionally that the size of the part of the sample from D is also fixed and equal to* $\#(S \cap D) = n_D$*. Then the BLUP assumes the form*

$$\widehat{Y}_D = \frac{N_D - n_D}{N_D}\frac{1}{n}\sum_{i \in S} Y_i + \frac{1}{N_D}\sum_{i \in S \cap D} Y_i \tag{10}$$

and its mean square error is

$$\mathbb{E}(\widehat{Y}_D - \bar{Y}_D)^2 = \frac{N_D - n_D}{nN_D^2}(N_D - n_D + n)\sigma^2. \tag{11}$$

Let us remark that the formula (10) is known; it can be found e.g. in Ghosh (2001, p. 58).

Proof Let us note that for a sampling plan with fixed sample size n, we have

$$\pi^T \mathbf{1} = n \tag{12}$$

and

$$P\mathbf{1} = n\pi, \tag{13}$$

see Chapter 2.6, Särndal et al. (1992).

(i) Note that $V\mathbf{1} = \sigma^2 \Pi \mathbf{1} + \mu^2(P\mathbf{1} - \pi\pi^T\mathbf{1})$. Thus by (12) and (13) we get $V\mathbf{1} = \sigma^2\pi$. Consequently, $V^{-1}\pi = \frac{1}{\sigma^2}\mathbf{1}$ and $\pi^T V^{-1}\pi = \frac{n}{\sigma^2}$. Now, (8) and (9) follow from (3) and (4).

(ii) From our additional assumption combined with (12) and (13) it follows that $\pi^T\mathbf{e}_D = n_D$ and $P\mathbf{e}_D = n_D\pi$. Thus $V\mathbf{e}_D = \sigma^2\Pi\mathbf{e}_D + \mu^2(P\mathbf{e}_D - \pi\pi^T\mathbf{e}_D) = \sigma^2\pi_D$. Consequently, $V^{-1}\pi_D = \frac{1}{\sigma^2}\mathbf{e}_D$ and $\pi_D^T V^{-1}\pi_D = \pi_D^T V^{-1}\pi = \frac{n_D}{\sigma^2}$. Now, (10) and (11) follow from (5) and (6). □

Remark 2 In the special case when $D = U$, formulas (10) and (11) reduce to

$$\widehat{Y}_U = \frac{1}{n}\sum_{i \in S} Y_i$$

and

$$\mathbb{E}(\widehat{Y}_U - \bar{Y}_U)^2 = \left(\frac{1}{n} - \frac{1}{N}\right)\sigma^2.$$

The predictor $\widehat{Y}_U$ from the above remark can be found e.g. in Scott and Smith (1969). In fact, their theorem is valid for a larger class of estimators than we consider, namely for estimators of the form $\sum_{i \in S} w_i(S)Y_i$ (i.e. linear in Y_is for each fixed sample S but not necessarily linear in Z_is). See also Cassel et al. (1977, Sect. 5.2, Remark 1 (iii)).

4 Optimal sampling strategy

We now tackle the problem of finding the optimal sampling strategies. An optimal strategy is a pair: an estimator/predictor and a sampling plan, which together minimize the mean square error. In Sect. 3 we derived the optimal estimators/predictors for an arbitrary but fixed sampling design (p). What is left, is to find optimal sampling plans (the second element of optimal strategies). We will show that an optimal strategy consists of any design with a fixed sample size together with the BLUE/BLUP - thus the strategies described in Corollary 1 in Sect. 3 are optimal. The main result of this section compares any sampling plan (p) to another plan $(\bar{p})$ with a fixed sample size(s) equal to expected sample size(s) for the plan (p). The proof is based on inequalities for positive definite matrices which are discussed in Appendix A.2.

Theorem 2 *Let (p) be a sampling design with the sample size $\#S = n$.*

(i) *Then*

$$Var(\hat{\mu}) \geq \frac{\sigma^2}{\mathbb{E}n}. \tag{14}$$

Consequently, if $(\bar{p})$ is a sampling design with the fixed sample size $\#\bar{S} = \bar{n} = \mathbb{E}n$ then the BLUE for $(\bar{p})$ has the variance less than or equal to that of the BLUE for (p).

(ii) *For (p) denote additionally $n_D = \#(S \cap D)$. Then*

$$\mathbb{E}\left(\widehat{Y}_D - \bar{Y}_D\right)^2 \geq \frac{(1-\frac{\mathbb{E}n_D}{N_D})^2\sigma^2}{\mathbb{E}n} - \frac{\sigma^2}{N_D^2}\mathbb{E}n_D + \frac{\sigma^2}{N_D}. \tag{15}$$

Consequently, if $(\bar{p})$ is a sampling design with the fixed sample sizes $\#\bar{S} = \bar{n} = \mathbb{E}n$ and $\#(\bar{S} \cap D) = \bar{n}_D = \mathbb{E}n_D$ then the BLUP for $(\bar{p})$ has the MSE less than or equal to that of the BLUP for (p).

Proof (i) The formula (14) follows from (4) and (25) upon substituting $K = V$, $B = \sigma^2\Pi$ and $\mathbf{v} = \pi$, taking into account that $\pi^T\mathbf{1} = \mathbb{E}n$ for any sampling design. Apparently, the right-hand side of (14) is the $(\bar{p})$-version of (9).

(ii) The proof of (15) is similar. We use (6) and (24) substituting $K = V$, $B = \sigma^2\Pi$, $\mathbf{v} = \pi$ and $\mathbf{u} = \frac{\sigma^2}{N_D}\pi_D$, taking into account that $\pi^T\mathbf{e}_D = \mathbb{E}n_D$. Again, note that the right-hand side of (15) is the $(\bar{p})$-version of (11). □

5 Symmetric designs

The general formulas for BLUE and BLUP, developed in Sect. 3, which are valid for arbitrary sampling plans, assume much simpler form for the class of symmetric designs.

For a symmetric design (p), that is a design which is invariant with respect to permutations of the population $U = \{1, \dots, N\}$, we have

$$\pi_i = \frac{\mathbb{E}n}{N} \quad \text{and} \quad \pi_{ij} = \frac{\mathbb{E}n(n-1)}{N(N-1)} \quad \text{for } i \neq j,\ i, j \in U. \tag{16}$$

Proposition 1 *Let (p) be a symmetric design. Then the best linear unbiased estimator (BLUE) $\hat{\mu}$ of μ has the form*

$$\hat{\mu} = \frac{1}{\mathbb{E}n} \sum_{i \in S} Y_i.$$

The variance of BLUE is the following

$$Var(\hat{\mu}) = \frac{\sigma^2}{\mathbb{E}n} + \mu^2 \frac{Var(n)}{(\mathbb{E}n)^2}.$$

Proof Note that

$$\pi = \frac{\mathbb{E}n}{N}\mathbf{1} \quad \text{and} \quad P = \left(\frac{\mathbb{E}n}{N} - \frac{\mathbb{E}n(n-1)}{N(N-1)} \right) I + \frac{\mathbb{E}n(n-1)}{N(N-1)} \mathbf{1}\mathbf{1}^T.$$

Consequently, in view of (1), we have

$$V = aI + b\mathbf{1}\mathbf{1}^T,$$

where

$$a = \sigma^2 \frac{\mathbb{E}n}{N} + \mu^2 \frac{\mathbb{E}n(N-n)}{N(N-1)} > 0 \quad \text{and} \quad b = \mu^2 \left(\frac{\mathbb{E}n(n-1)}{N(N-1)} - \left(\frac{\mathbb{E}n}{N} \right)^2 \right).$$

Using the well-known formula

$$(A + \mathbf{b}\mathbf{b}^T)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{b}\mathbf{b}^T A^{-1}}{1 + \mathbf{b}^T A^{-1} \mathbf{b}} \tag{17}$$

we get

$$V^{-1} = \frac{1}{a} \left(I - \frac{b}{a + Nb} \mathbf{1}\mathbf{1}^T \right).$$

Note that $a + Nb = [\sigma^2 \mathbb{E}n + \mu^2 Var(n)]/N > 0$. By Theorem 1 we get the result. □

Now let us turn to the best linear unbiased predictor (BLUP) $\widehat{Y}_D$ of the mean $\bar{Y}_D = \frac{1}{N_D}\mathbf{Y}^T \mathbf{e}_D$ in a domain $D \subset U$. In this case it makes sense to consider designs invariant with respect to all such permutations of the population U for which sets D

and $U \setminus D$ are invariant. Now the counterpart of the formula (16) is the following

$$\pi_i = \begin{cases} \frac{\mathbb{E}n_D}{N_D} & \text{for } i \in D, \\ \frac{\mathbb{E}n - \mathbb{E}n_D}{N - N_D} & \text{for } i \notin D, \end{cases}$$

$$\pi_{ij} = \begin{cases} \frac{\mathbb{E}n_D(n_D - 1)}{N_D(N_D - 1)} & \text{for } i \neq j,\ i, j \in D, \\ \frac{\mathbb{E}n_D(n - n_D)}{N_D(N - N_D)} & \text{for } i \in D,\ j \notin D \text{ or } i \notin D,\ j \in D, \\ \frac{\mathbb{E}(n - n_D)(n - n_D - 1)}{(N - N_D)(N - N_D - 1)} & \text{for } i \neq j,\ i, j \notin D. \end{cases} \tag{18}$$

Proposition 2 *Let* (p) *be a design satisfying* (18). *Then the best linear unbiased predictor (BLUP)* $\widehat{Y}_D$ *of* $\bar{Y}_D$ *has the form*

$$\widehat{Y}_D = \frac{\alpha}{\alpha + \beta} \cdot \frac{1}{\mathbb{E}n_D} \sum_{i \in S \cap D} Y_i + \frac{\beta}{\alpha + \beta} \cdot \frac{1}{\mathbb{E}n} \sum_{i \in S} Y_i, \tag{19}$$

where

$$\alpha = \frac{\mathbb{E}n_D}{N_D} \left(\frac{1}{\mathbb{E}n_D} - \frac{1}{\mathbb{E}n} \right) + \frac{1}{\gamma^2 \mathbb{E}n} \left(\frac{Var(n)}{\mathbb{E}n} - \frac{Cov(n, n_D)}{\mathbb{E}n_D} \right),$$

and

$$\beta = \frac{N_D - \mathbb{E}n_D}{N_D} \left(\frac{1}{\mathbb{E}n_D} - \frac{1}{\mathbb{E}n} \right) + \frac{1}{\gamma^2 \mathbb{E}n_D} \left(\frac{Var(n_D)}{\mathbb{E}n_D} - \frac{Cov(n, n_D)}{\mathbb{E}n} \right).$$

The MSE of the BLUP is the following

$$\mathbb{E}\left[(\widehat{Y}_D - \bar{Y}_D)^2 \right] = \sigma^2 \left[\frac{N_D - \mathbb{E}n_D}{N_D \mathbb{E}n} + \frac{Cov(n, n_D)}{\gamma^2 \mathbb{E}n \mathbb{E}n_D} + \frac{\alpha\beta}{\alpha + \beta} \right]. \tag{20}$$

Proof Let us introduce the following notations: $E = U \setminus D$, $N_E = N - N_D$, $n_E = n - n_D$. Let $\mathbf{1}_D$ and $\mathbf{1}_E$ be the N_D-dimensional and N_E-dimensional vectors of 1s, respectively. Denote by I_D, I_E the identity matrices indexed by the elements of D and E, respectively (we can assume without loss of generality that units in D precede those in E). Inserting the formulas preceding the formulation of this proposition into (1) we obtain the matrix V with the following block structure:

$$V = \begin{pmatrix} a_D I_D + b_D \mathbf{1}_D \mathbf{1}_D^T & c\, \mathbf{1}_D \mathbf{1}_E^T \\ c\, \mathbf{1}_E \mathbf{1}_D^T & a_E I_E + b_E \mathbf{1}_E \mathbf{1}_E^T \end{pmatrix},$$

where

$$a_D = \sigma^2 \frac{\mathbb{E}n_D}{N_D} + \mu^2 \frac{\mathbb{E}n_D(N_D - n_D)}{N_D(N_D - 1)}, \quad b_D = \mu^2 \left(\frac{\mathbb{E}n_D(n_D - 1)}{N_D(N_D - 1)} - \left[\frac{\mathbb{E}n_D}{N_D} \right]^2 \right),$$

$$a_E = \sigma^2 \frac{\mathbb{E}n_E}{N_E} + \mu^2 \frac{\mathbb{E}n_E(N_E - n_E)}{N_E(N_E - 1)}, \quad b_E = \mu^2 \left(\frac{\mathbb{E}n_E(n_E - 1)}{N_E(N_E - 1)} - \left[\frac{\mathbb{E}n_E}{N_E} \right]^2 \right),$$

$$c = \mu^2 \frac{Cov(n_D, n_E)}{N_D N_E}.$$

To compute V^{-1} one can use the well-known formula for inverting a block matrix:

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix}^{-1} = \begin{pmatrix} (A - BC^{-1}B^T)^{-1} & -A^{-1}B(C - B^T A^{-1} B)^{-1} \\ -(C - B^T A^{-1} B)^{-1} B^T A^{-1} & (C - B^T A^{-1} B)^{-1} \end{pmatrix}.$$

Together with (17) it gives

$$V^{-1} = \frac{1}{\Delta} \begin{pmatrix} \frac{\Delta I_D - [b_D(a_E + b_E N_E) - c^2 N_E] \mathbf{1}_D \mathbf{1}_D^T}{a_D} & -c\, \mathbf{1}_D \mathbf{1}_E^T \\ -c\, \mathbf{1}_E \mathbf{1}_D^T & \frac{\Delta I_E - [b_E(a_D + b_D N_D) - c^2 N_D] \mathbf{1}_E \mathbf{1}_E^T}{a_E} \end{pmatrix},$$

where $\Delta = (a_D + b_D N_D)(a_E + b_E N_E) - c^2 N_D N_E$. Then, in view of (2), (5) and (6), after elementary but tedious computation, we arrive at the desired results. □

Remark 3 Let us note that for $D = U$, from the limiting cases of formulas (19) and (20), we obtain

$$\widehat{Y}_U = \frac{1}{\mathbb{E}n} \sum_{i \in S} Y_i \quad \text{and} \quad \mathbb{E}\left(\widehat{Y}_U - \bar{Y}_U\right)^2 = \sigma^2 \left(\frac{1}{\mathbb{E}n} - \frac{1}{N} + \frac{Var(n)}{\gamma^2 (\mathbb{E}n)^2} \right).$$

Appendix A

A.1 Minimization

Our main tool in Sect. 3 is the following general result on constrained minimization.

Lemma 1 *Let X be a random variable, $\mathbf{Y}$ a random vector in $\mathbb{R}^N$ and $\mathbf{r} \in \mathbb{R}^N$ be such that $\mathbf{r}\mathbb{E}X = \mathbb{E}\mathbf{Y}$. Denote by Σ the covariance matrix of $\mathbf{Y}$, by $\mathbf{c}$ – the (column) vector of covariances of X and $\mathbf{Y}$ and by σ^2 – the variance of X. Then*

$$\inf_{\mathbf{w} \in \mathbb{R}^N : \mathbf{w}^T \mathbf{r} = 1} \mathbb{E}(\mathbf{w}^T \mathbf{Y} - X)^2 = \mathbb{E}(\mathbf{w}_{\text{opt}}^T \mathbf{Y} - X)^2 = \frac{(1 - \mathbf{c}^T \Sigma^{-1} \mathbf{r})^2}{\mathbf{r}^T \Sigma^{-1} \mathbf{r}} - \mathbf{c}^T \Sigma^{-1} \mathbf{c} + \sigma^2 \tag{21}$$

and

$$\mathbf{w}_{\text{opt}} = \frac{1 - \mathbf{c}^T \Sigma^{-1} \mathbf{r}}{\mathbf{r}^T \Sigma^{-1} \mathbf{r}} \Sigma^{-1} \mathbf{r} + \Sigma^{-1} \mathbf{c}. \tag{22}$$

Proof Let us define families of random variables X_μ and random vectors $\mathbf{Y}_\mu$ as follows

$$X_\mu = X - \mathbb{E}X + \mu \quad \text{and} \quad \mathbf{Y}_\mu = \mathbf{Y} - \mathbf{r}\mathbb{E}X + \mathbf{r}\mu$$

indexed by a parameter $\mu \in \mathbb{R}$. Let us observe that

$$\mathbf{Y}_\mu = \mathbf{r}\mu + \frac{\mathbf{c}}{\sigma^2}(X_\mu - \mu) + \left(\mathbf{Y}_\mu - \frac{\mathbf{c}}{\sigma^2} X_\mu + \frac{\mathbf{c}}{\sigma^2}\mu - \mathbf{r}\mu\right). \tag{23}$$

Note that (23) defines a mixed linear model with scalar fixed effect μ and random effect $X_\mu - \mu$. One can easily see that the random effect is uncorrelated with the error term $\mathbf{Y}_\mu - \frac{\mathbf{c}}{\sigma^2} X_\mu + \frac{\mathbf{c}}{\sigma^2}\mu - \mathbf{r}\mu$ because of the definition of $\mathbf{c}$. Since $\mathbf{w}^T \mathbf{Y}_\mu - X_\mu = \mathbf{w}^T \mathbf{Y} - X$ for every μ, our minimization problem is equivalent to finding the best unbiased predictor of $X_\mu = \mu + (X_\mu - \mu)$ (the constraint $\mathbf{w}^T \mathbf{r} = 1$ amounts to the requirement of unbiasedness). The standard formula for the best predictor in a mixed linear model (see for example Rao (2003, Chap. 6), gives the result. □

Let us mention that Lemma 1 can also be proved directly by a standard Lagrange multiplier method, without any reference to linear models.

A.2 Inequalities for positive definite matrices

To prove Theorem 2 on the optimal strategy in Sect. 4, we need the following auxiliary results on matrices.

Lemma 2 *If $K = B + C$ where B and C are symmetric $N \times N$ matrices, B is positive definite and C is nonzero nonnegative definite, then $M = B^{-1} - K^{-1}$ is positive definite.*

Proof Note that $M = B^{-1}(K - B)K^{-1} = B^{-1} C K^{-1}$. Therefore for any $\mathbf{u} \in \mathbb{R}^N$

$$\begin{aligned} \mathbf{u}^T M \mathbf{u} = \mathbf{u}^T B^{-1} C K^{-1} \mathbf{u} &= \mathbf{v}^T K B^{-1} C \mathbf{v} = \mathbf{v}^T (B + C) B^{-1} C \mathbf{v} \\ &= \mathbf{v}^T B B^{-1} C \mathbf{v} + \mathbf{v}^T C B^{-1} C \mathbf{v} \\ &= \mathbf{v}^T C \mathbf{v} + \mathbf{w}^T B^{-1} \mathbf{w}, \end{aligned}$$

with $\mathbf{v} = K^{-1}\mathbf{u}$ and $\mathbf{w} = C\mathbf{v}$. Now if $\mathbf{u} \neq \mathbf{0}$ then $\mathbf{u}^T M \mathbf{u} > 0$ since $\mathbf{w}^T B^{-1} \mathbf{w} > 0$ and $\mathbf{v}^T C \mathbf{v} \geq 0$. □

Proposition 3 *Let $K = B + C$ where B and C are symmetric $N \times N$ matrices, B is positive definite and C is nonzero nonnegative definite. Then for any non-zero vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$*

$$\frac{(1 - \mathbf{u}^T K^{-1} \mathbf{v})^2}{\mathbf{v}^T K^{-1} \mathbf{v}} - \mathbf{u}^T K^{-1} \mathbf{u} \geq \frac{(1 - \mathbf{u}^T B^{-1} \mathbf{v})^2}{\mathbf{v}^T B^{-1} \mathbf{v}} - \mathbf{u}^T B^{-1} \mathbf{u} \tag{24}$$

Remark 4 Note that for any $\mathbf{v} \in \mathbb{R}^N$,

$$\frac{1}{\mathbf{v}^T K^{-1} \mathbf{v}} \geq \frac{1}{\mathbf{v}^T B^{-1} \mathbf{v}}. \tag{25}$$

This follows by taking $\mathbf{u} = \alpha \mathbf{v}$ for a real number α in (24).

Proof of Proposition 3 Let $M = B^{-1} - K^{-1}$. Then $\mathbf{v}^T K^{-1} \mathbf{v} = a - \alpha$, $\mathbf{u}^T K^{-1} \mathbf{v} = b - \beta$ and $\mathbf{u}^T K^{-1} \mathbf{u} = c - \gamma$, where $a = \mathbf{v}^T B^{-1} \mathbf{v}$, $b = \mathbf{u}^T B^{-1} \mathbf{v}$, $c = \mathbf{u}^T B^{-1} \mathbf{u}$, $\alpha = \mathbf{v}^T M \mathbf{v}$, $\beta = \mathbf{u}^T M \mathbf{v}$ and $\gamma = \mathbf{u}^T M \mathbf{u}$. Denote additionally $d = 1 - b$. Then (24) can be rewritten as

$$\frac{(d + \beta)^2}{a - \alpha} + \gamma \geq \frac{d^2}{a}. \tag{26}$$

Note that $a > 0$ since B^{-1} is positive definite and $a - \alpha > 0$ since K^{-1} is positive definite. Therefore (26) is equivalent to

$$L = (d + \beta)^2 a + a(a - \alpha)\gamma - d^2(a - \alpha) \geq 0. \tag{27}$$

To prove that $L \geq 0$ we consider separately two cases: (1) $d^2 < a\gamma$, (2) $d^2 \geq a\gamma$.

In the case (1) the inequality (27) holds since then $L > (d + \beta)^2 a \geq 0$.

In the case (2) we first note that if $d = 0$ then (27) holds since then $L = a\left(\beta^2 + (a - \alpha)\gamma\right) > 0$. Note that due to Lemma 2 the matrix M is positive definite and thus $\gamma > 0$. If $d \neq 0$ we write the inequality (27) as

$$H(x) := d^2(\gamma x^2 + 2\beta x + \alpha) \geq a(\alpha\gamma - \beta^2), \tag{28}$$

where $x = a/d$. Note that the function H is quadratic with the positive coefficient γd^2 of x^2. Then it has the minimum

$$H_0 = d^2 \frac{\alpha\gamma - \beta^2}{\gamma}. \tag{29}$$

To show that the inequality (28) holds it suffices to prove that

$$H_0 \geq a(\alpha\gamma - \beta^2).$$

Again we use Lemma 2: positive definiteness of M allows to use the Cauchy–Schwartz inequality for the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T M \mathbf{v}$. Consequently $\alpha\gamma \geq \beta^2$ and either both sides of (29) are zero or it takes the form $d^2 \geq a\gamma$, which is trivially satisfied in the considered case. □

References

Bellhouse DR (1984) A review of optimal designs in survey sampling. Can J Stat 12(1):53–65

Cassel C, Särndal C-E, Wretman J (1977) Foundations of Inference in Survey Sampling. Wiley, New York

Chambers R (2005) An introduction to model-based survey sampling. In: Krapavickaitė D, Plikusas A (eds) Workshop on survey sampling theory and methodology. Statistics Lithuania, Vilnius, pp 9–51

Ghosh M (2001) Model-dependent small area estimation—theory and practice. In: Lehtonen R, Djerf K (eds) Lecture notes on estimation for population domains and small areas. Statistics Finland, Helsinki, pp 51–108

Gonzalez ME (1973) Use and evaluation of synthetic estimates. In: Proceedings of the social statistics section. American Mathematical Association, pp 33–36

Marker DA (1999) Organization of small area estimators using a generalized linear regression framework. J Off Stat 15(1):1–24

Royall RM (1970) On finite population sampling theory under certain linear regression models. Biometrika 57(2):377–387

Rao JNK (2003) Small area estimation. Wiley-Interscience, New York

Scott A, Smith TMF (1969) A note on estimating secondary characteristics in multivariate surveys. Sankhyã Ser A 31:497–499

Särndal C-E, Swensson B, Wretman J (1992) Model assisted survey sampling. Springer, New York