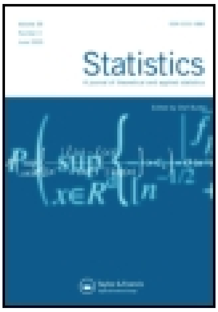

Statistics: A Journal of Theoretical and Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gsta20

Linear estimation and prediction under model-design approach with small area effects

Wojciech Niemiro [a] [b] & Jacek Wesołowski [b] [c]

[a] Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland

[b] Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland

[c] Central Statistical Office, Warsaw, Poland

Available online: 23 Feb 2011

Statistics, iFirst, 2011, 1–25

Taylor & Francis
Taylor & Francis Group

Linear estimation and prediction under model-design approach with small area effects

Wojciech Niemiro[a,b] and Jacek Wesołowski[b,c]*

[a] *Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland;* [b] *Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland;* [c] *Central Statistical Office, Warsaw, Poland*

(*Received 3 December 2007; final version received 8 September 2010*)

We consider a problem of small area estimation under a mixed linear model with area-specific auxiliary variables and random area effects. The model design approach is consistently pursued. The mathematical results of this paper support the view (not generally accepted in the literature) that while using a super-population model, even under non-informative sampling, the sample selection process cannot be completely ignored. Estimators and predictors under consideration are linear in observations, that is in variables Z_i's, where Z_i is the value of the variable of interest if i is in the sample and zero otherwise. This notion of linearity is different from that prevailing in the literature. Unbiasedness and optimality are understood with respect to both the model and the sampling design. We consider general sampling designs, for which sample sizes in small areas can be random. We show that the best linear unbiased estimators (BLUEs) and best linear unbiased predictors (BLUPs) in general do not exist. However, they do exist if the sample sizes in small areas are fixed. Moreover, we prove that such designs are optimal. Empirical versions (EBLUE and EBLUPs) are also derived and numerically tested. In simulation experiments, we examine the mean square error of estimates/predictors, the coverage rates of confidence intervals and the predictive power of auxiliary variables. Rather unexpectedly, the proposed predictors turn out to be quite robust against model misspecification. A special case of Bernoulli sampling is examined in detail as an illustrative example.

Keywords: super-population; BLUP; linear model; model design approach; small area estimation

MSC 2000: 62D05; 62J05

1. Introduction

Model assumptions appear more and more frequently in the theory of sample surveys. This is especially visible in small area estimation, where basically it is the only way which may lead to the valid inference for domains with very small sample sizes. Since the paper by Fay and Herriot [1], several types of linear models have been considered in this area. The monograph by Rao [2] gives a thorough review and is the basic reference for this theory. An earlier survey paper by Marker [3] gives a unified view of small area estimation through a general linear regression framework. An excellent description of estimation techniques in this field is given in an even earlier paper by Ghosh and Rao [4]. Mostly, as soon as the model assumptions are imposed,

*Corresponding author. Email: J.Wesolowski@mini.pw.edu.pl

ISSN 0233-1888 print/ISSN 1029-4910 online

DOI: 10.1080/02331888.2010.542656
http://www.informaworld.com

design considerations are left aside (see, for instance, [5] or [6]). That is, the conditional approach is adopted where conditioning is with respect to the sample. Many authors considered such a purely model-based approach. Standard references in this setting are several papers by Royall and co-authors starting with Royall [7]. A review can be found in [8, Chapter 5]. For more recent results, one can also consult Bellhouse [9] or Chandra and Chambers [10]. The rule which seems to prevail in the literature is made explicit in an important review paper by Pfeffermann [11] who writes 'when the selection probabilities are not related to the values of the response variable, the models holding for the population hold also for the sample data and the sample selection process can be ignored' (par. 6). While in many cases, this conclusion is justified, it may not be correct in general, even under non-informative sampling. This phenomenon is related to random sample sizes. Such a situation occurs when one is interested in optimality with respect to model and design criteria jointly. This has already been observed in our earlier paper [12], where synthetic and composite estimators under the model design approach were obtained in a very simple model with no small area effect. In particular, in that paper, we gave an improvement of the estimator obtained in [13]. Another direction of research in small area estimation is to minimize the role of model assumptions. This can be achieved through a careful choice of the survey design and through efficient direct domain estimators as developed, e.g. in [14,15] or [16].

The main novelty of our approach lies in examining rigorously and consistently the influence of *both the model and the design* on the inference. In this sense, the present paper continues the research of Niemiro and Wesołowski [12]. The focus is on theoretical aspects of the model design approach to small area estimation. We consider a mixed linear model with fixed effects, auxiliary variables (which are small-area-specific), random small area effects and random unit effects (errors) – this is for the model part of our setup. The sample is chosen according to a given design plan and only sampled units are observed – this is for the design part of our setup. We consider non-informative but otherwise entirely arbitrary sampling plans. This approach is described thoroughly in Section 2. In Section 3, properties of the covariance matrix of observations are investigated. In Section 4, we study the existence of best linear unbiased estimator (BLUE) and two best linear unbiased predictors (BLUPs) under such super-population assumptions. Multivariate auxiliary variables known at the level of small areas are considered. In particular, we show through Theorem 1 (see Remark 2) that for a wide range of designs, BLUE and BLUP do not exist. However, they do exist for designs with fixed sample sizes. In Section 5, we prove that the optimal strategy for BLUE and BLUPs is to apply any design with fixed sample sizes. Such a situation happens in the case of stratified sampling when small areas are unions of several strata (for instance, in the survey of small enterprises conducted by the Central Statistical Office in Poland – see Example 1 in Section 7). The case of univariate auxiliary variables, essentially different from the case of multivariate auxiliary variable, is considered in Section 6. Empirical versions of the BLUPs are derived in Section 7. Theoretical results are accompanied by three numerical experiments. We examine the mean square error (MSE) of the estimates/predictors, the coverage rates of confidence intervals and the predictive power of auxiliary variables. The first experiment assumes a model without auxiliary variables; the second one with one auxiliary variable; in the third, the inference is based on an incorrectly specified model. In Section 8, our general results on BLUE and BLUPs, in the case of univariate auxiliary variable, are applied to the simplest sampling plan with random sample sizes, the Bernoulli sampling. We also give an account of a small-scale simulation study for this case.

2. Setup

By $U = \{1, \dots, N\}$ we denote a population. The population is partitioned into M disjoint small areas $(U_m)_{m=1,\dots,M}$, i.e. $U = \bigcup_{m=1}^{M} U_m$. Without any loss of generality, we will assume that the ordering of the population is such that $U_m = \{i_m, i_m + 1, \ldots, i_{m+1} - 1\}$, $m = 1, \ldots, M$, with $1 = i_1 < i_2 < \cdots < i_M < i_{M+1} = N + 1$. Then $i_{m+1} - i_m = N_m$ denotes the number of units in U_m, $m = 1, \ldots, M$. For any small area U_m, a non-random vector of auxiliary variables $\underline{x}_m = (x_{m,1}, \ldots, x_{m,q})^{\mathrm{T}}$ is given, $m = 1, \ldots, M$.

With each element $i \in U$, we associate a random variable Y_i. We assume that the random vector $\underline{Y} = (Y_1, \ldots, Y_N)^{\mathrm{T}}$ has the following structure:

$$\underline{Y} = \boldsymbol{e}(\mathbb{X}\underline{\beta} + \underline{u}) + \underline{\varepsilon}, \tag{1}$$

where $\boldsymbol{e}$ is an $N \times M$ matrix whose mth column is $\underline{e}_m$ with the ith component equal to 1 if $i \in U_m$ and otherwise equal to 0, $m = 1, \ldots, M$, $\mathbb{X}$ is an $M \times q$-dimensional matrix with $\underline{x}_m^{\mathrm{T}}$ being its mth row, $\underline{\beta} = (\beta_1, \ldots, \beta_q)$ is an unknown vector of parameters, $\underline{u} = (u_1, \ldots, u_M)^{\mathrm{T}}$ is a random vector such that $\mathbb{E}(\underline{u}) = \underline{0}$ and $\mathrm{Cov}(\underline{u}) = v^2\mathbb{I}_M$ and $\underline{\varepsilon} = (\varepsilon_1, \ldots \varepsilon_N)^{\mathrm{T}}$ is a random vector such that $\mathbb{E}(\underline{\varepsilon}) = \underline{0}$ and $\mathrm{Cov}(\underline{\varepsilon}) = \sigma^2\mathbb{I}_N$ with $\sigma^2 > 0$, where $\mathbb{I}_k$ stands for the $k \times k$ identity matrix, $k = M, N$. Although the homoscedasticity assumption, which we adopted here, is not always a realistic assumption, it eliminates serious difficulties related to the estimation of unit-specific variances. Therefore, the homoscedastic approach is fairly standard in the literature (see, e.g. [17]). Moreover, we assume that $\underline{u}$ and $\underline{\varepsilon}$ are independent.

Alternatively,

$$Y_i = \underline{x}_m^{\mathrm{T}}\underline{\beta} + u_m + \varepsilon_i$$

for any $i \in U_m$, $m = 1, \ldots, M$. Let us stress that auxiliary variables are assumed to be known at the small area level (and not at the unit level). This assumption is essential for the theoretical results to be presented. In practice, it may be justified in situations when the unit-specific auxiliary information is unavailable.

The sampling design p is a distribution of indicators of elements being sampled, i.e. $P(\underline{I} = \underline{\delta}) = p(\underline{\delta})$, where $\underline{I} = (I_1, \ldots, I_n)^{\mathrm{T}}$, with $I_i = 1$ if the ith element is chosen to the sample, otherwise it is 0, $i \in U$, and $\underline{\delta} = (\delta_1, \ldots, \delta_N) \in \{0, 1\}^N$. With p we associate a vector $\underline{\pi} = (\pi_1, \ldots, \pi_N)^{\mathrm{T}}$ of inclusion probabilities of the first order, i.e. $\pi_i = P(I_i = 1)$, $i \in U$, a related diagonal matrix $\Pi = \mathrm{diag}(\underline{\pi})$ and an $N \times N$ matrix $\mathbb{P} = [\pi_{ij}]$ of inclusion probabilities of the second order, i.e. $\pi_{ij} = P(I_i = 1, I_j = 1)$, $i, j \in U$. Throughout this paper, we assume that $\underline{I}$ and $\underline{Y}$ are independent, that is, the sampling plan p is non-informative.

Additionally, we introduce an $N \times M$ matrix $\tilde{\pi}$ whose mth column $\underline{\pi}_m$ is a vector having the ith component equal to π_i if $i \in U_m$ and 0 if $i \notin U_m$, $m = 1, \ldots, M$, and an $N \times N$ matrix $\tilde{\mathbb{P}}$ which is a block-diagonal matrix, with the mth diagonal block $\tilde{\mathbb{P}}_m$ associated with the mth small area in the sense that $\tilde{\mathbb{P}}_m = [\pi_{ij}]_{i,j \in U_m}$, i.e. it is a matrix of second-order inclusion probabilities for the restriction of p to the set U_m, $m = 1, \ldots, M$.

Let $\underline{Z} = (Z_1, \ldots, Z_N) = \mathrm{diag}(\underline{I})\underline{Y}$ be the vector of observations, i.e. $Z_i = Y_i I_i$, $i \in U$.

Note that $\mathbb{E}\underline{Y} = \boldsymbol{e}\mathbb{X}\underline{\beta}$ and

$$\mathbb{E}\underline{Z} = \tilde{\pi}\mathbb{X}\underline{\beta} = \sum_{m=1}^{M} x_m^{\mathrm{T}}\underline{\beta}\underline{\pi}_m.$$

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 to both the sampling plan p and the super-population model $\underline{Y}$. Thus for a function f,

$$\mathbb{E}f(\underline{Y}, \underline{I}) = \int_{\mathbf{R}^N} \sum_{\underline{\delta} \in \{0,1\}^N} f(\mathbf{y}, \underline{\delta})p(\underline{\delta})P_{\underline{Y}}(\mathrm{d}\mathbf{y}),$$

where $P_{\underline{Y}}$ denotes the probability distribution of the random vector $\underline{Y}$. Note that in some papers, frequently cited in the literature (e.g. [17, Chapter 12]), 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.

Moreover, on noting that $\text{diag}(\underline{I})\mathbb{e}\mathbb{X}\underline{\beta} = \text{diag}(\mathbb{e}\mathbb{X}\underline{\beta})\underline{I}$, we obtain the expression for the covariance matrix of vector $\underline{Z}$ of observations, $\mathbb{K} = \text{Cov}\,(\underline{Z})$ as

$$\mathbb{K} = \sigma^2 \mathbb{\Pi} + \text{diag}(\mathbb{e}\mathbb{X}\underline{\beta})(\mathbb{P} - \underline{\pi}\underline{\pi}^{\mathrm{T}})\text{diag}(\mathbb{e}\mathbb{X}\underline{\beta}) + v^2\tilde{\mathbb{P}}.$$

3. Properties of the covariance matrix $\mathbb{K}$

Note that the covariance matrix $\mathbb{K}$ can be decomposed as

$$\mathbb{K} = \mathbb{B} + \mathbb{D} \tag{2}$$

with

$$\mathbb{B} = \text{diag}(\mathbb{e}\mathbb{X}\underline{\beta})(\mathbb{P} - \underline{\pi}\underline{\pi}^{\mathrm{T}})\text{diag}(\mathbb{e}\mathbb{X}\underline{\beta}) + v^2(\tilde{\mathbb{P}} - \tilde{\pi}\tilde{\pi}^{\mathrm{T}})$$

and

$$\mathbb{D} = \sigma^2 \mathbb{\Pi} + v^2 \tilde{\pi}\tilde{\pi}^{\mathrm{T}}.$$

Matrix $\tilde{\mathbb{P}} - \tilde{\pi}\tilde{\pi}^{\mathrm{T}}$ is a non-negative definite since its mth diagonal block is a covariance matrix of a subvector of indicators $(I_i, i \in U_m)$, $m = 1, \ldots, M$, and thus $\mathbb{B}$ is a non-negative definite matrix, while $\mathbb{D}$ is a positive definite matrix under the assumption that $\sigma^2 > 0$.

Alternatively, $\mathbb{D}$ can be represented in the block-diagonal form as

$$\mathbb{D} = \text{Diag}(\sigma^2 \mathbb{\Pi}_m + v^2 \underline{\pi}_m \underline{\pi}_m^{\mathrm{T}}),$$

where $\mathbb{\Pi}_m$ is the diagonal block of the diagonal matrix $\mathbb{\Pi}$ representing the small area U_m, $m = 1, \ldots, M$. This representation is convenient for obtaining the form of $\mathbb{D}^{-1}$ by block-wise inversions

$$\mathbb{D}^{-1} = \text{Diag}[(\sigma^2 \mathbb{\Pi}_m + v^2 \underline{\pi}_m \underline{\pi}_m^{\mathrm{T}})^{-1}].$$

Thus, using the well-known formula for the inversion of $\mathbb{A} + \underline{b}\underline{b}^{\mathrm{T}}$, where $\mathbb{A}$ is an invertible matrix and $\underline{b}$ is a vector, we get

$$\mathbb{D}^{-1} = \sigma^{-2}\text{Diag}\left(\mathbb{\Pi}_m^{-1} - \frac{v^2 \underline{e}_m \underline{e}_m^{\mathrm{T}}}{\sigma^2 + v^2 \mathbb{E}(n_m)}\right),$$

where n_m is the number of the elements of the sample belonging to U_m and thus $\mathbb{E}(n_m) = \underline{\pi}_m^{\mathrm{T}}\underline{1} = \sum_{i \in U_m} \pi_i$, $m = 1, \ldots, M$. Note that

$$\mathbb{D}^{-1}\underline{\pi}_m = \frac{\underline{e}_m}{\sigma^2 + v^2 \mathbb{E}(n_m)}, \quad m = 1, \ldots, M, \tag{3}$$

and, consequently, since $\underline{\pi} = \sum_{m=1}^{M} \underline{\pi}_m$, we get

$$\mathbb{D}^{-1}\underline{\pi} = \sum_{m=1}^{M} \frac{\underline{e}_m}{\sigma^2 + v^2 \mathbb{E}(n_m)}.$$

Moreover, since

$$\tilde{\pi}\mathbb{X} = \sum_{m=1}^{M} \underline{\pi}_m \underline{x}_m^{\mathrm{T}} \tag{4}$$

by Equation (3)

$$\mathbb{D}^{-1}\tilde{\pi}\mathbb{X} = \sum_{m=1}^{M} \mathbb{D}^{-1}\underline{\pi}_m \underline{x}_m^{\mathrm{T}} = \sum_{m=1}^{M} \frac{\underline{e}_m \underline{x}_m^{\mathrm{T}}}{\sigma^2 + v^2 \mathbb{E} n_m} \tag{5}$$

and due to the relation $\underline{\pi}_m^{\mathrm{T}} \underline{e}_m = \mathbb{E} n_m$

$$\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\mathbb{D}^{-1}\tilde{\pi}\mathbb{X} = \mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}} \sum_{m=1}^{M} \frac{\underline{e}_m \underline{x}_m^{\mathrm{T}}}{\sigma^2 + v^2 \mathbb{E} n_m} = \sum_{m=1}^{M} \frac{\underline{x}_m \underline{x}_m^{\mathrm{T}} \mathbb{E} n_m}{\sigma^2 + v^2 \mathbb{E} n_m}. \tag{6}$$

Note that for sampling plans with fixed sample sizes $(n_m)_{m=1,\ldots,M}$ in small areas $(U_m)_{m=1,\ldots,M}$, we have

$$(\mathbb{P} - \underline{\pi}\underline{\pi}^{\mathrm{T}})\underline{e}_m = \underline{0} \quad \text{and} \quad (\tilde{\mathbb{P}} - \underline{\tilde{\pi}}\underline{\tilde{\pi}}^{\mathrm{T}})\underline{e}_m = \underline{0} \tag{7}$$

for any $m = 1, \ldots, M$.

Consequently, for any $m = 1, \ldots, M$,

$$\mathbb{K}\underline{e}_m = \mathbb{D}\underline{e}_m = (\sigma^2 + v^2 n_m)\underline{\pi}_m,$$

which yields

$$\mathbb{K}^{-1}\underline{\pi}_m = \frac{\underline{e}_m}{\sigma^2 + v^2 n_m}, \quad m = 1, \ldots, M. \tag{8}$$

Consequently, in this case, the respective formulas are very similar to the ones for the matrix $\mathbb{D}$:

$$\mathbb{K}^{-1}\tilde{\pi}\mathbb{X} = \sum_{m=1}^{M} \frac{\underline{e}_m \underline{x}_m^{\mathrm{T}}}{\sigma^2 + v^2 n_m} \tag{9}$$

and

$$\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\mathbb{K}^{-1}\tilde{\pi}\mathbb{X} = \sum_{m=1}^{M} \frac{n_m \underline{x}_m \underline{x}_m^{\mathrm{T}}}{\sigma^2 + v^2 n_m}. \tag{10}$$

4. Multivariate auxiliary variables

We start with an auxiliary minimization result which will be our main tool in this section.

LEMMA 1 *Let $\underline{X}$ be a random vector in $\mathbf{R}^q$ and $\underline{W}$ be a random vector in $\mathbf{R}^N$. Let $\underline{\gamma}$ be a fixed vector in $\mathbf{R}^q$ and $\mathbb{A}$ be an $N \times q$ matrix satisfying $\mathbb{E}\underline{W} = \mathbb{A}\mathbb{E}\underline{X}$. Denote $\boldsymbol{\Sigma} = \mathrm{Cov}\,(\underline{W})$, $\mathbb{C} = \mathrm{Cov}\,(\underline{W}, \underline{X}) = \mathbb{E}[(\underline{W} - \mathbb{E}\underline{W})(\underline{X} - \mathbb{E}\underline{X})^{\mathrm{T}}]$ which is an $N \times q$ matrix and $\mathbb{S} = \mathrm{Cov}\,(\underline{X})$. Then*

$$\inf_{\underline{w}\in\mathbf{R}^N : \mathbb{A}^{\mathrm{T}}\underline{w} = \underline{\gamma}} \mathbb{E}(\underline{w}^{\mathrm{T}}\underline{W} - \underline{\gamma}^{\mathrm{T}}\underline{X})^2 = \mathbb{E}(\underline{w}_{\mathrm{opt}}^{\mathrm{T}}\underline{W} - \underline{\gamma}^{\mathrm{T}}\underline{X})^2$$
$$= \underline{\gamma}^{\mathrm{T}}(\mathbb{I}_q - \mathbb{C}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbb{A})(\mathbb{A}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbb{A})^{-1}(\mathbb{I}_q - \mathbb{A}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbb{C})\underline{\gamma} - \underline{\gamma}^{\mathrm{T}}\mathbb{C}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbb{C}\underline{\gamma} + \underline{\gamma}^{\mathrm{T}}\mathbb{S}\underline{\gamma} \tag{11}$$

and

$$\underline{w}_{\mathrm{opt}} = \boldsymbol{\Sigma}^{-1}\mathbb{A}(\mathbb{A}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbb{A})^{-1}(\mathbb{I}_q - \mathbb{A}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbb{C})\underline{\gamma} + \boldsymbol{\Sigma}^{-1}\mathbb{C}\underline{\gamma}. \tag{12}$$

Proof Using the Lagrange method, we will minimize the function

$$F(\underline{w}) = \mathbb{E}(\underline{w}^{\mathrm{T}}\underline{W} - \underline{\gamma}^{\mathrm{T}}\underline{X})^2 - 2\underline{\lambda}^{\mathrm{T}}\mathbb{A}^{\mathrm{T}}\underline{w}.$$

Note that its derivative, DF is of the form

$$(\mathrm{DF})(\underline{w}) = 2\mathbb{E}[(\underline{w}^{\mathrm{T}}\underline{W} - \underline{\gamma}^{\mathrm{T}}\underline{X})\underline{W}] - 2\mathbb{A}\underline{\lambda}.$$

Thus, we arrive at the equation

$$[\mathbb{E}\underline{W}\underline{W}^{\mathrm{T}}]\underline{w} - [\mathbb{E}\underline{W}\underline{X}^{\mathrm{T}}]\underline{\gamma} = \mathbb{A}\underline{\lambda},$$

which is equivalent to

$$\mathbf{\Sigma}\underline{w} + (\mathbb{E}\underline{W})(\mathbb{E}\underline{W})^{\mathrm{T}}\underline{w} - \mathbb{C}\underline{\gamma} - (\mathbb{E}\underline{W})(\mathbb{E}\underline{X})^{\mathrm{T}}\underline{\gamma} = \mathbb{A}\underline{\lambda}.$$

Now using the assumption $\mathbb{E}\underline{W} = \mathbb{A}\mathbb{E}\underline{X}$ and the constraint $\mathbb{A}^{\mathrm{T}}\underline{w} = \underline{\gamma}$, we arrive at the equation

$$\mathbf{\Sigma}\underline{w} - \mathbb{C}\underline{\gamma} = \mathbb{A}\underline{\lambda}$$

and thus

$$\underline{w} = \mathbf{\Sigma}^{-1}\mathbb{A}\underline{\lambda} + \mathbf{\Sigma}^{-1}\mathbb{C}\underline{\gamma}.$$

Using again the constraint, we obtain

$$\underline{\lambda} = (\mathbb{A}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbb{A})^{-1}\underline{\gamma} - (\mathbb{A}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbb{A})^{-1}\mathbb{A}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbb{C}\underline{\gamma}$$

which leads to Equation (12) and consequently to Equation (11). ■

In the result below, we consider linear combinations of the observations $(Z_i)_{i\in U}$ with coefficients which may depend on v^2, σ^2 and $\underline{\beta}$. Such a function can be regarded as a linear estimator or predictor if its coefficients depend only on v^2 and σ^2 (assumed to be known).

THEOREM 1 *Consider the model as described in Section* 2. *By D we denote a fixed small area, i.e.* $D \in \{U_m, : m = 1, \ldots, M\}$.

(1) *Let*

$$\underline{\tilde{\beta}} = (\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\mathbb{K}^{-1}\tilde{\pi}\mathbb{X})^{-1}\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\mathbb{K}^{-1}\underline{Z}. \tag{13}$$

Then, $\mathbb{E}\underline{\tilde{\beta}} = \underline{\beta}$ *and the covariance matrix of* $\underline{\tilde{\beta}}$ *is given by*

$$\mathrm{Cov}\,\underline{\tilde{\beta}} = (\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\mathbb{K}^{-1}\tilde{\pi}\mathbb{X})^{-1}. \tag{14}$$

Let $\underline{\hat{\beta}}$ *be an arbitrary linear unbiased estimator of* $\underline{\beta}$. *Then for any* $\underline{\gamma} \in \mathbf{R}^q$,

$$\mathrm{Var}\,\underline{\gamma}^{\mathrm{T}}\underline{\hat{\beta}} \geq \mathrm{Var}\,\underline{\gamma}^{\mathrm{T}}\underline{\tilde{\beta}}. \tag{15}$$

(2) *Let* $\theta_D = \underline{x}_D^{\mathrm{T}}\underline{\beta} + u_D$ *and*

$$\tilde{\theta}_D = (x_D - v^2\underline{\pi}_D^{\mathrm{T}}\mathbb{K}^{-1}\tilde{\pi}\mathbb{X})\underline{\tilde{\beta}} + v^2\underline{\pi}_D^{\mathrm{T}}\mathbb{K}^{-1}\underline{Z}. \tag{16}$$

Then, $\mathbb{E}\tilde{\theta}_D = \theta_D$ *and*

$$\begin{aligned}\mathbb{E}(\tilde{\theta}_D - \theta_D)^2 = v^2 - v^4\underline{\pi}_D^{\mathrm{T}}\mathbb{K}^{-1}\underline{\pi}_D + (\underline{x}_D^{\mathrm{T}} - v^2\underline{\pi}_D^{\mathrm{T}}\mathbb{K}^{-1}\tilde{\pi}\mathbb{X})(\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\mathbb{K}^{-1}\tilde{\pi}\mathbb{X})^{-1}\\ \times (\underline{x}_D - v^2\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\mathbb{K}^{-1}\underline{\pi}_D).\end{aligned} \tag{17}$$

For an arbitrary linear unbiased predictor $\hat{\theta}_D$ *of* θ_D, *its MSE satisfies*

$$\mathbb{E}(\hat{\theta}_D - \theta_D)^2 \geq \mathbb{E}(\tilde{\theta}_D - \theta_D)^2. \tag{18}$$

(3) *Let* $\bar{Y}_D = \underline{x}_D^{\mathrm{T}}\underline{\beta} + u_D + \bar{\varepsilon}_D$, *where* $\bar{\varepsilon}_D = (1/N_D)\sum_{i\in D}\varepsilon_i$. *Let*

$$\tilde{Y}_D = \left(\underline{x}_D^{\mathrm{T}} - \left(v^2 + \frac{\sigma^2}{N_D}\right)\underline{\pi}_D^{\mathrm{T}}\mathbb{K}^{-1}\tilde{\pi}\mathbb{X}\right)\hat{\beta} + \left(v^2 + \frac{\sigma^2}{N_D}\right)\underline{\pi}_D^{\mathrm{T}}\mathbb{K}^{-1}\underline{Z}. \tag{19}$$

Then, $\mathbb{E}\tilde{Y}_D = \bar{Y}_D$ *and*

$$\mathbb{E}(\tilde{Y}_D - \bar{Y}_D)^2 = v^2 + \frac{\sigma^2}{N_D} - \left(v^2 + \frac{\sigma^2}{N_D}\right)^2\underline{\pi}_D^{\mathrm{T}}\mathbb{K}^{-1}\underline{\pi}_D + \left(\underline{x}_D^{\mathrm{T}} - \left(v^2 + \frac{\sigma^2}{N_D}\right)\underline{\pi}_D^{\mathrm{T}}\mathbb{K}^{-1}\tilde{\pi}\mathbb{X}\right)$$
$$\times\,(\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\mathbb{K}^{-1}\tilde{\pi}\mathbb{X})^{-1}\left(\underline{x}_D - \left(v^2 + \frac{\sigma^2}{N_D}\right)\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\mathbb{K}^{-1}\underline{\pi}_D\right). \tag{20}$$

For an arbitrary linear unbiased predictor $\hat{Y}_D$ *of* $\bar{Y}_D$, *its MSE satisfies*

$$\mathbb{E}(\hat{Y}_D - \bar{Y}_D)^2 \ge \mathbb{E}(\tilde{Y}_D - \bar{Y}_D)^2. \tag{21}$$

Remark 1 Note that $\underline{\tilde{\beta}}$ may be regarded as the best linear unbiased *pseudo*-estimator of $\underline{\beta}$, $\tilde{\theta}_D$ as the best linear unbiased *pseudo*-predictor of θ_D and $\tilde{Y}_D$ as the best linear unbiased *pseudo*-predictor of $\bar{Y}_D$. The prefix *pseudo* is necessary because all these random variables depend, in general, on the unknown quantity $\underline{\beta}$ through $\mathbb{K}$.

Proof of Theorem 1 In each of three parts, the proof is based on Lemma 1 with $\underline{W} = \underline{Z}$, $\mathbf{\Sigma} = \mathbb{K}$, $\mathbb{A} = \tilde{\pi}\mathbb{X}$ common for all these parts. Other quantities in each of these parts differ.

Part 1. We apply Lemma 1 additionally assuming $\underline{X} = \underline{\beta}$ (thus $\underline{X}$ is here non-random and consequently, $\mathbb{S} = \mathbb{0}$, and $\mathbb{C} = \mathbb{0}$). Now it follows from Equation (12) that for any $\underline{\gamma} \in \mathbf{R}^q$

$$\underline{w}_{\mathrm{opt}}^{\mathrm{T}}\underline{W} = \underline{w}_{\mathrm{opt}}^{\mathrm{T}}(\underline{\beta})\underline{W} = \underline{\gamma}^{\mathrm{T}}\underline{\tilde{\beta}},$$

where $\underline{\tilde{\beta}}$ is defined by Equation (13). Moreover, since the second part of Equation (11) holds for any $\underline{\gamma} \in \mathbf{R}^q$, the covariance of $\underline{\tilde{\beta}}$ is given by Equation (14).

For any $\underline{\gamma} \in \mathbf{R}^q$ consider a linear unbiased estimator $\underline{\gamma}^{\mathrm{T}}\underline{\hat{\beta}} = \underline{w}^{\mathrm{T}}\underline{Z}$ of $\underline{\gamma}^{\mathrm{T}}\underline{\beta}$ and note that the unbiasedness condition is equivalent to $\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\underline{w} = \underline{\gamma}$. Therefore, Lemma 1 implies Inequality (15), that is,

$$\mathbb{E}(\underline{\gamma}^{\mathrm{T}}\underline{\hat{\beta}} - \underline{\gamma}^{\mathrm{T}}\underline{\beta})^2 \ge \mathbb{E}(\underline{\gamma}^{\mathrm{T}}\underline{\tilde{\beta}} - \underline{\gamma}^{\mathrm{T}}\underline{\beta})^2 = \underline{\gamma}^{\mathrm{T}}(\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\mathbb{K}^{-1}\tilde{\pi}\mathbb{X})^{-1}\underline{\gamma}.$$

Part 2. The proof repeats the scheme of the previous one. The only difference is that in Lemma 1 we use

$$\underline{X} = \underline{\beta} + u_D\frac{\underline{x}_D}{\underline{x}_D^{\mathrm{T}}\underline{x}_D}, \quad \text{and thus} \quad \mathbb{S} = \frac{v^2}{\underline{x}_D^{\mathrm{T}}\underline{x}_D}\underline{x}_D\underline{x}_D^{\mathrm{T}},$$
$$\mathbb{C} = v^2\frac{\underline{\pi}_D\underline{x}_D^{\mathrm{T}}}{\underline{x}_D^{\mathrm{T}}\underline{x}_D} \quad \text{and} \quad \underline{\gamma} = \underline{x}_D.$$

Upon inserting the above quantities into Equations (12) and (11), after a little of algebra, we arrive at Equations (16) and (17). Again Inequality (18) follows from Lemma 1 just as in Part 1.

Part 3. Again we use the same scheme. Now in Lemma 1, we use additionally

$$\underline{X} = \underline{\beta} + (u_D + \bar{\varepsilon}_D)\frac{\underline{x}_D}{\underline{x}_D^{\mathrm{T}}\underline{x}_D}, \quad \text{and thus} \quad \mathbb{S} = \left(v^2 + \frac{\sigma^2}{N_D}\right)\frac{\underline{x}_D\underline{x}_D^{\mathrm{T}}}{\underline{x}_D^{\mathrm{T}}\underline{x}_D},$$

$$\mathbb{C} = \left(v^2 + \frac{\sigma^2}{N_D}\right)\frac{\underline{\pi}_D\underline{x}_D^{\mathrm{T}}}{\underline{x}_D^{\mathrm{T}}\underline{x}_D} \quad \text{and} \quad \underline{\gamma} = \underline{x}_D.$$

Similarly, as earlier Equations (12) and (11), after a little of algebra lead to Equations (19) and (20). Again Inequality (21) is a direct consequence of Lemma 1. ■

Remark 2 Expanding further Remark 1, we observe that, in general, the covariance of $\underline{\tilde{\beta}}$ and the MSEs of $\tilde{\theta}_D$ and $\tilde{Y}_D$ as given in Equations (14), (17) and (20) depend on unknown $\underline{\beta}$. As long as such dependence takes place, even if variances v^2 and σ^2 are known: (i) the BLUE of $\underline{\beta}$ does not exist; (ii) the BLUP of θ_D does not exist; (iii) the BLUP of $\bar{Y}_D$ does not exist. It follows from the inequalities in Theorem 1 and the fact that w_{opt} in Lemma 1 is unique.

Note that if

$$(\mathbb{P} - \underline{\pi}\underline{\pi}^{\mathrm{T}})\mathbb{e} = \mathbb{0} \tag{22}$$

then the dependence on $\underline{\beta}$ in the formulas of Theorem 1 vanishes. The most interesting case when it occurs is when the sample sizes are fixed. This situation is considered in the next result.

COROLLARY 1 *Let p be a sample plan with fixed sample sizes $(n_m)_{m=1,\ldots,M}$ in small areas $(U_m)_{m=1,\ldots,M}$. Then*

(1) *BLUE $\underline{\tilde{\beta}}$ of $\underline{\beta}$ has the form*

$$\underline{\tilde{\beta}} = \left(\sum_{m=1}^{M}\kappa_m\underline{x}_m\underline{x}_m^{\mathrm{T}}\right)^{-1}\sum_{m=1}^{M}\kappa_m\underline{x}_m\tilde{Z}_m, \tag{23}$$

where $\kappa_m = v^2 n_m/(\sigma^2 + v^2 n_m)$, $\tilde{Z}_m = (1/n_m)\sum_{i\in U_m} Z_i$, $m = 1,\ldots,M$.

The covariance matrix of $\underline{\tilde{\beta}}$ is

$$\operatorname{Cov}\underline{\tilde{\beta}} = v^2\left(\sum_{m=1}^{M}\kappa_m\underline{x}_m\underline{x}_m^{\mathrm{T}}\right)^{-1}. \tag{24}$$

(2) *BLUP $\tilde{\theta}_D$ of θ_D has the form*

$$\tilde{\theta}_D = (1-\kappa_D)\underline{x}_D^{\mathrm{T}}\underline{\tilde{\beta}} + \kappa_D\tilde{Z}_D \tag{25}$$

with $\underline{\tilde{\beta}}$ given by Equation (23), and its MSE is

$$\mathbb{E}(\tilde{\theta}_D - \theta_D)^2 = v^2(1-\kappa_D)\left[(1-\kappa_D)\underline{x}_D^{\mathrm{T}}\left(\sum_{m=1}^{M}\kappa_m\underline{x}_m\underline{x}_m^{\mathrm{T}}\right)^{-1}\underline{x}_D + 1\right]. \tag{26}$$

(3) *BLUP $\tilde{Y}_D$ of $\bar{Y}_D$ has the form*

$$\tilde{Y}_D = (1 - \kappa_D \tau_D)\underline{x}_D^{\mathrm{T}}\underline{\tilde{\beta}} + \kappa_D \tau_D \tilde{Z}_D \tag{27}$$

with $\tau_D = 1 + (\sigma^2/v^2 N_D)$ and $\underline{\tilde{\beta}}$ given by Equation (23). *Its MSE is*

$$\mathbb{E}(\tilde{Y}_D - \bar{Y}_D)^2 = v^2(1 - \kappa_D \tau_D)\left[(1 - \kappa_D \tau_D)\underline{x}_D^{\mathrm{T}}\left(\sum_{m=1}^{M}\kappa_m \underline{x}_m \underline{x}_m^{\mathrm{T}}\right)^{-1}\underline{x}_D + \tau_D\right]. \tag{28}$$

Proof To begin with, note that for sample plans with fixed sample sizes $(n_m)_{m=1,\dots,M}$ condition (7) is satisfied and consequently $\mathbb{K}$ does not depend on $\underline{\beta}$.

Part 1. The result given in Equation (23) follows directly from Equation (13) by referring to Equations (9) and (10), while Equation (24) follows directly from Equation (14) by referring to Equation (10) alone.

Part 2. The result given in Equation (25) follows from Equation (16) by referring to Equations (8) and (4). Now Equation (26) follows from Equation (17) by referring to Equations (10) and (8), and Equation (4) again.

Part 3. The result given in Equation (27) follows from Equation (19) through Equations (8) and (4). Similarly, Equation (28) is a consequence of Equation (20) due to Equations (10) and (8) combined with Equation (4). ■

5. Optimal strategy

For the proof of the optimality theorem given later in this section, we first need the following result on positive definiteness of some matrices.

PROPOSITION 1 *Let $\boldsymbol{\Sigma} = \mathbb{B} + \mathbb{D}$, where $\mathbb{B}$ and $\mathbb{D}$ are symmetric $N \times N$ matrices, $\mathbb{B}$ is non-negative definite and $\mathbb{D}$ is positive definite. Then for any $N \times q$ $(q \leq N)$ matrices $\mathbb{A}$ and $\mathbb{C}$, such that $\mathbb{A}$ is of full rank, matrix*

$$\begin{aligned}(\mathbb{I}_q - \mathbb{C}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbb{A})(\mathbb{A}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbb{A})^{-1}(\mathbb{I}_q - \mathbb{A}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbb{C}) - \mathbb{C}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbb{C}\\ - [(\mathbb{I}_q - \mathbb{C}^{\mathrm{T}}\mathbb{D}^{-1}\mathbb{A})(\mathbb{A}^{\mathrm{T}}\mathbb{D}^{-1}\mathbb{A})^{-1}(\mathbb{I}_q - \mathbb{A}^{\mathrm{T}}\mathbb{D}^{-1}\mathbb{C}) - \mathbb{C}^{\mathrm{T}}\mathbb{D}^{-1}\mathbb{C}]\end{aligned}$$

is positive definite.

Proof Let $\mathbb{V} = \boldsymbol{\Sigma}^{-1}$ and $\mathbb{V} + \mathbb{H} = \mathbb{D}^{-1}$. Note that $\mathbb{H}$ is positive definite. For $t \in [0, 1]$, we define $\mathbb{V}_t = \mathbb{V} + t\mathbb{H}$. For an arbitrary fixed $\underline{\lambda} \in \mathbf{R}^q$, we consider function $f : [0, 1] \to \mathbf{R}$ given by

$$f(t) = \underline{\lambda}^{\mathrm{T}}[(\mathbb{I}_q - \mathbb{C}^{\mathrm{T}}\mathbb{V}_t\mathbb{A})(\mathbb{A}^{\mathrm{T}}\mathbb{V}_t\mathbb{A})^{-1}(\mathbb{I}_q - \mathbb{A}^{\mathrm{T}}\mathbb{V}_t\mathbb{C}) - \mathbb{C}^{\mathrm{T}}\mathbb{V}_t\mathbb{C}]\underline{\lambda}.$$

Since $\underline{\lambda}$ is arbitrary to finish the proof, it is sufficient to show that $f(0) > f(1)$. To this end, we compute the derivative f'

$$\begin{aligned}f'(t) = &-\underline{\lambda}^{\mathrm{T}}\mathbb{C}^{\mathrm{T}}\mathbb{H}\mathbb{A}(\mathbb{A}^{\mathrm{T}}\mathbb{V}_t\mathbb{A})^{-1}(\mathbb{I}_q - \mathbb{A}^{\mathrm{T}}\mathbb{V}_t\mathbb{C})\underline{\lambda}\\ &- \underline{\lambda}^{\mathrm{T}}(\mathbb{I}_q - \mathbb{C}^{\mathrm{T}}\mathbb{V}_t\mathbb{A})(\mathbb{A}^{\mathrm{T}}\mathbb{V}_t\mathbb{A})^{-1}\mathbb{A}^{\mathrm{T}}\mathbb{H}\mathbb{A}(\mathbb{A}^{\mathrm{T}}\mathbb{V}_t\mathbb{A})^{-1}(\mathbb{I}_q - \mathbb{A}^{\mathrm{T}}\mathbb{V}_t\mathbb{C})\underline{\lambda}\\ &- \underline{\lambda}^{\mathrm{T}}(\mathbb{I}_q - \mathbb{C}^{\mathrm{T}}\mathbb{V}_t\mathbb{A})(\mathbb{A}^{\mathrm{T}}\mathbb{V}_t\mathbb{A})^{-1}\mathbb{A}^{\mathrm{T}}\mathbb{H}\mathbb{C}\underline{\lambda}\\ &- \underline{\lambda}^{\mathrm{T}}\mathbb{C}^{\mathrm{T}}\mathbb{H}\mathbb{C}\underline{\lambda}\\ = &-\underline{\lambda}^{\mathrm{T}}(\mathbb{C}^{\mathrm{T}} + \mathbb{Z}_t^{\mathrm{T}})\mathbb{H}(\mathbb{C} + \mathbb{Z}_t)\underline{\lambda},\end{aligned}$$

where $\mathbb{Z}_t = \mathbb{A}(\mathbb{A}^{\mathrm{T}}\mathbb{V}_t\mathbb{A})^{-1}(\mathbb{I}_q - \mathbb{A}^{\mathrm{T}}\mathbb{V}_t\mathbb{C})$. Thus, f' is negative and consequently f is decreasing. ■

THEOREM 2 *Let p be an arbitrary sampling design. For p we denote sample sizes in small areas by $n_m = \#(S \cap U_m)$, $m = 1, \ldots, M$. Let $\bar{p}$ be a sampling design with fixed sample sizes $\#(\bar{S} \cap U_m) = \bar{n}_m$, $m = 1, \ldots, M$. Assume that $\bar{n}_m = \mathbb{E}n_m$, $m = 1, \ldots, M$.*

(1) *For any $\underline{\gamma} \in \mathbf{R}^q$ and any linear unbiased estimator $\hat{\underline{\beta}}$ of $\underline{\beta}$*

$$\operatorname{Var} \underline{\gamma}^{\mathrm{T}} \hat{\underline{\beta}} \geq \operatorname{Var} \underline{\gamma}^{\mathrm{T}} \tilde{\underline{\beta}} \geq \underline{\gamma}^{\mathrm{T}} \left(\sum_{m=1}^{M} \frac{\underline{x}_m \underline{x}_m^{\mathrm{T}} \mathbb{E}n_m}{\sigma^2 + v^2 \mathbb{E}n_m} \right)^{-1} \underline{\gamma}, \tag{29}$$

where $\tilde{\underline{\beta}}$ is given by Equation (13). *The above inequality means that for any linear functional $\underline{\gamma}^{\mathrm{T}}\underline{\beta}$ of $\underline{\beta}$, its BLUE under $\bar{p}$ has a variance less than or equal to the variance of any linear unbiased estimator $\underline{\gamma}^{\mathrm{T}}\hat{\underline{\beta}}$ under p.*

(2) *The MSE of any linear unbiased predictor $\hat{\theta}_D$ of θ_D satisfies*

$$\begin{aligned} \mathbb{E}(\hat{\theta}_D - \theta_D)^2 &\geq \mathbb{E}(\tilde{\theta}_D - \theta_D)^2 \\ &\geq \frac{\sigma^2}{\sigma^2 + v^2 \mathbb{E}n_D} \left(\frac{\sigma^2}{\sigma^2 + v^2 \mathbb{E}n_D} \underline{x}_D^{\mathrm{T}} \left(\sum_{m=1}^{M} \frac{\underline{x}_m \underline{x}_m^{\mathrm{T}} \mathbb{E}n_m}{\sigma^2 + v^2 \mathbb{E}n_m} \right)^{-1} \underline{x}_D + v^2 \right), \end{aligned} \tag{30}$$

where $\tilde{\theta}_D$ is given by Equation (16). *The above inequality means that the BLUP of θ_D for sampling design $\bar{p}$ has the MSE less than or equal to the MSE of any linear unbiased predictor $\hat{\theta}_D$ for p.*

(3) *The MSE of any linear unbiased predictor $\hat{Y}_D$ of $\bar{Y}_D$ satisfies*

$$\begin{aligned} \mathbb{E}(\hat{Y}_D - \bar{Y}_D)^2 \geq \mathbb{E}(\tilde{Y}_D - \bar{Y}_D)^2 \geq \frac{\sigma^2(N_D - \mathbb{E}n_D)}{N_D(\sigma^2 + v^2\mathbb{E}n_D)} \\ \cdot \left(\frac{\sigma^2(N_D - \mathbb{E}n_D)}{N_D(\sigma^2 + v^2\mathbb{E}n_D)} \underline{x}_D^{\mathrm{T}} \left(\sum_{m=1}^{M} \frac{\underline{x}_m \underline{x}_m^{\mathrm{T}} \mathbb{E}n_m}{\sigma^2 + v^2 \mathbb{E}n_m} \right)^{-1} \underline{x}_D + v^2 + \frac{\sigma^2}{N_D} \right), \end{aligned} \tag{31}$$

where $\tilde{Y}_D$ is given by Equation (19). *The above inequality means that the BLUP of $\bar{Y}_D$ for sampling design $\bar{p}$ has the MSE less than or equal to the MSE of any linear unbiased predictor $\hat{Y}_D$ for p.*

Proof In all cases, we will use Proposition 1 with $\mathbf{\Sigma} = \mathbb{K}$, matrices $\mathbb{B}$ and $\mathbb{D}$ defined through the decomposition of $\mathbb{K}$ given in Equation (2) and $\mathbb{A} = \tilde{\pi}\mathbb{X}$. Only matrix $\mathbb{C}$ of Proposition 1 will be chosen in a different way in each case below.

To prove the first part, we take $\mathbb{C} = \mathbb{0}$. Therefore,

$$\operatorname{Var} \underline{\gamma}^{\mathrm{T}} \tilde{\underline{\beta}} = \underline{\gamma}^{\mathrm{T}} (\mathbb{X}^{\mathrm{T}} \tilde{\pi}^{\mathrm{T}} \mathbb{K}^{-1} \tilde{\pi} \mathbb{X})^{-1} \underline{\gamma} \geq \underline{\gamma}^{\mathrm{T}} (\mathbb{X}^{\mathrm{T}} \tilde{\pi}^{\mathrm{T}} \mathbb{D}^{-1} \tilde{\pi} \mathbb{X})^{-1} \underline{\gamma}$$

and thus the Formula (29) follows from Equation (6). The conclusion follows from the first part of Corollary 1.

To prove the second part, we take $\mathbb{C} = (v^2/\underline{x}_D^{\mathrm{T}}\underline{x}_D)\underline{\pi}_D\underline{x}_D^{\mathrm{T}}$. Thus

$$\mathbb{E}(\tilde{\theta}_D - \theta_D)^2 \geq (\underline{x}_D^{\mathrm{T}} - v^2\underline{\pi}_D^{\mathrm{T}}\mathbb{D}^{-1}\tilde{\pi}\mathbb{X})(\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\mathbb{D}^{-1}\tilde{\pi}\mathbb{X})^{-1}(\underline{x}_D - v^2\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\mathbb{D}^{-1}\underline{\pi}_D) - v^4\underline{\pi}_D^{\mathrm{T}}\mathbb{D}^{-1}\underline{\pi}_D + v^2$$

and thus the Formula (30) follows from Equations (3), (4) and (6). The conclusion follows from the second part of Corollary 1.

To prove the third part, we take $\mathbb{C} = (v^2 + \sigma^2/N_D)\underline{\pi}_D\underline{x}_D^{\mathrm{T}}/\underline{x}_D^{\mathrm{T}}\underline{x}_D$. Thus

$$\mathbb{E}(\tilde{Y}_D - \bar{Y}_D)^2 \geq v^2 + \frac{\sigma^2}{N_D} - \left(v^2 + \frac{\sigma^2}{N_D}\right)^2 \underline{\pi}_D^{\mathrm{T}}\mathbb{D}^{-1}\underline{\pi}_D + \left(\underline{x}_D^{\mathrm{T}} - \left(v^2 + \frac{\sigma^2}{N_D}\right)\underline{\pi}_D^{\mathrm{T}}\mathbb{D}^{-1}\tilde{\pi}\mathbb{X}\right) \times (\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\mathbb{D}^{-1}\tilde{\pi}\mathbb{X})^{-1}\left(\underline{x}_D - \left(v^2 + \frac{\sigma^2}{N_D}\right)\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\mathbb{D}^{-1}\underline{\pi}_D\right)$$

and again the Formula (31) follows from Equations (3), (4) and (6). The conclusion follows from the third part of Corollary 1. ∎

6. Univariate auxiliary variables

The situation which was considered in Section 4 becomes much simpler when variables x_m are one dimensional, i.e. $q = 1$. Let us begin with some new notation. Symbols x_m and β (not underlined) now denote scalars. Vector $\underline{x} = (x_1, \ldots, x_M)^{\mathrm{T}}$ takes over the role of matrix $\mathbb{X}$ and the basic model becomes

$$\underline{Y} = \mathbb{e}(\beta\underline{x} + \underline{u}) + \underline{\varepsilon}. \tag{32}$$

Other notation remains the same. Just as in Section 4, we have $\mathbb{E}(\underline{u}) = \underline{0}$, Cov $(\underline{u}) = v^2\mathbb{I}_M$, $\mathbb{E}(\underline{\varepsilon}) = \underline{0}$ and Cov $(\underline{\varepsilon}) = \sigma^2\mathbb{I}_N$ with $\sigma^2 > 0$. In this section, we will assume that parameters $\gamma^2 = (v^2 + \sigma^2)/\beta^2$ and $\tau^2 = \sigma^2/v^2$ are known. This might be the case for repeated surveys, when τ and γ could be estimated with relatively small error from the past. Note that γ is the coefficient of variation of $Y_i = x_m\beta + u_m + \varepsilon_i$ up to multiple x_m and τ is the ratio of standard deviations of unit error ε_i and small area effect u_m. We will show that in this case, the BLUE of β and the BLUPs of θ_D and $\bar{Y}_D$ exist, in contrast with Corollary 2. The situation when γ and/or τ are not known will be considered later, in Section 7. Let us now introduce a scaled version of $\mathbb{K}$ which is *known* under our present assumptions:

$$\bar{\mathbb{K}} = \frac{\mathbb{K}}{v^2} = \tau^2\Pi + \frac{1+\tau^2}{\gamma^2}\mathrm{diag}(\mathbb{e}\underline{x})(\mathbb{P} - \underline{\pi}\underline{\pi}^{\mathrm{T}})\mathrm{diag}(\mathbb{e}\underline{x}) + \tilde{\mathbb{P}}.$$

The following result is an immediate consequence of Theorem 1.

Proposition 2 *Consider the model as described in Section* 2 *assuming that* $q = 1$ *and that* $\gamma^2 = (v^2 + \sigma^2)/\beta^2$ *and* $\tau^2 = \sigma^2/v^2$ *are known.*

(1) *BLUE* $\hat{\beta}$ *of* β *has the form*

$$\hat{\beta} = \frac{\underline{x}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\bar{\mathbb{K}}^{-1}\underline{Z}}{\underline{x}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\bar{\mathbb{K}}^{-1}\tilde{\pi}\underline{x}} \tag{33}$$

and its variance is

$$\mathrm{Var}\,\hat{\beta} = \frac{v^2}{\underline{x}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\bar{\mathbb{K}}^{-1}\tilde{\pi}\underline{x}}. \tag{34}$$

(2) *BLUP* $\hat{\theta}_D$ *of* $\theta_D = x_D\beta + u_D$ *has the form*

$$\hat{\theta}_D = (x_D - \underline{x}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\bar{\mathbb{K}}^{-1}\underline{\pi}_D)\hat{\beta} + \underline{\pi}_D^{\mathrm{T}}\bar{\mathbb{K}}^{-1}\underline{Z} \tag{35}$$

and its MSE is

$$\mathbb{E}(\hat{\theta}_D - \theta_D)^2 = v^2\left[\frac{\left(x_D - \underline{x}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\bar{\mathbb{K}}^{-1}\underline{\pi}_D\right)^2}{\underline{x}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\bar{\mathbb{K}}^{-1}\tilde{\pi}\underline{x}} + 1 - \underline{\pi}_D^{\mathrm{T}}\bar{\mathbb{K}}^{-1}\underline{\pi}_D\right]. \tag{36}$$

(3) *BLUP* $\hat{Y}_D$ *of* $\bar{Y}_D = \theta_D + \bar{\varepsilon}_D$ *has the form*

$$\hat{Y}_D = \left[x_D - \tau_D\underline{x}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\bar{\mathbb{K}}^{-1}\underline{\pi}_D\right]\hat{\beta} + \tau_D\underline{\pi}_D^{\mathrm{T}}\bar{\mathbb{K}}^{-1}\underline{Z} \tag{37}$$

and its MSE is

$$\mathbb{E}(\hat{Y}_D - \bar{Y}_D)^2 = v^2\left[\tau_D - \tau_D^2\underline{\pi}_D^{\mathrm{T}}\bar{\mathbb{K}}^{-1}\underline{\pi}_D + \frac{\left(x_D - \tau_D\underline{x}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\bar{\mathbb{K}}^{-1}\underline{\pi}_D\right)^2}{\underline{x}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\bar{\mathbb{K}}^{-1}\tilde{\pi}\underline{x}}\right] \tag{38}$$

for $\tau_D = 1 + \tau^2/N_D$.

Proof It is sufficient to note that *pseudo*-estimator $\underline{\tilde{\beta}}$ and *pseudo*-predictors $\tilde{\theta}_D$ and $\tilde{Y}_D$ considered in Theorem 1 now become a genuine estimator (denoted by $\underline{\hat{\beta}}$) and predictors (denoted by $\hat{\theta}_D$ and $\hat{Y}_D$), because they are free from unknown parameters. ■

The formulas for the variance and MSEs derived above can be used to construct approximate confidence intervals. Unknown parameters in these formulas should then be replaced by their consistent estimators. A similar, rather standard, plug-in approach will be used in Section 7 to derive empirical versions of BLUE and BLUPs.

COROLLARY 2 *Let* p *be a sample plan with fixed sample sizes* $(n_m)_{m=1,\dots,M}$ *in small areas* $(U_m)_{m=1,\dots,M}$. *Then*

(1) *BLUE* $\hat{\beta}$ *of* β *has the form*

$$\hat{\beta} = \frac{\sum_{m=1}^{M}\kappa_m x_m\tilde{Z}_m}{\sum_{m=1}^{M}\kappa_m x_m^2}, \tag{39}$$

where $\kappa_m = n_m/(\tau^2 + n_m)$, $\tilde{Z}_m = (1/n_m)\sum_{i\in U_m} Z_i$.

The variance of $\hat{\beta}$ *is*

$$\operatorname{Var}\hat{\beta} = \frac{v^2}{\sum_{m=1}^{M}\kappa_m x_m^2}. \tag{40}$$

(2) *BLUP* $\hat{\theta}_D$ *of* θ_D *has the form*

$$\hat{\theta}_D = (1-\kappa_D)x_D\hat{\beta} + \kappa_D\tilde{Z}_D \tag{41}$$

with $\hat{\beta}$ *given by Equation* (39), *and its MSE is*

$$\mathbb{E}(\hat{\theta}_D - \theta_D)^2 = v^2(1-\kappa_D)\left[\frac{(1-\kappa_D)x_D^2}{\sum_{m=1}^{M}\kappa_m x_m^2} + 1\right]. \tag{42}$$

(3) *BLUP $\hat{Y}_D$ of $\bar{Y}_D$ has the form*

$$\hat{Y}_D = (1 - \kappa_D \tau_D) x_D \hat{\beta} + \kappa_D \tau_D \tilde{Z}_D \tag{43}$$

with $\hat{\beta}$ given by Equation (39), *and its MSE is*

$$\mathbb{E}(\hat{Y}_D - \bar{Y}_D)^2 = v^2 (1 - \kappa_D \tau_D) \left[\frac{(1 - \kappa_D \tau_D) x_D^2}{\sum_{m=1}^{M} \kappa_m x_m^2} + \tau_D \right], \tag{44}$$

where $\tau_D = 1 + \tau^2 / N_D$.

7. EBLUE and EBLUPs

Parameters v^2 and σ^2 are usually unknown in practice. To use results of previous sections, one has to replace these variance components by their estimates. We will also show that $\underline{\beta}$ can be simultaneously estimated. This allows us to transform pseudo-estimator $\underline{\tilde{\beta}}$ and pseudo-predictors $\tilde{\theta}_D$ and $\tilde{Y}_D$ into computable estimator EBLUE and predictors EBLUPs.

Recall that

$$\begin{aligned} Z_i &= I_i (\underline{x}_m^{\mathrm{T}} \underline{\beta} + u_m + \varepsilon_i), \quad i \in U_m, \\ \operatorname{Var} Z_i &= \mathbb{E}(Z_i - \pi_i \underline{x}_m^{\mathrm{T}} \underline{\beta})^2 = [\sigma^2 + v^2 + (\underline{x}_m^{\mathrm{T}} \underline{\beta})^2 (1 - \pi_i)] \pi_i, \\ \operatorname{Cov}(Z_i, Z_j) &= \begin{cases} v^2 \pi_{ij} + (\underline{x}_m^{\mathrm{T}} \underline{\beta})^2 [\pi_{ij} - \pi_i \pi_j] & \text{for } i, j \in U_m; \\ (\underline{x}_{m_1}^{\mathrm{T}} \underline{\beta})(\underline{x}_{m_2}^{\mathrm{T}} \underline{\beta}) [\pi_{ij} - \pi_i \pi_j] & \text{for } i \in U_{m_1}, j \in U_{m_2}. \end{cases} \end{aligned}$$

Note that

$$\sum_{m=1}^{M} \mathbb{E} \left(\frac{I(n_m > 1)}{n_m - 1} \sum_{i \in S \cap U_m} (Z_i - \tilde{Z}_m)^2 \right) = \sigma^2 \sum_{m=1}^{M} \Pr(n_m > 1),$$

where

$$\tilde{Z}_m = \begin{cases} \frac{1}{n_m} \sum_{j \in U_m} Z_j & \text{if } n_m \neq 0; \\ 0 & \text{if } n_m = 0. \end{cases}$$

Thus, the method of moments leads to the following estimator of σ^2:

$$\check{\sigma}^2 = \frac{1}{\sum_{m=1}^{M} \Pr(n_m > 1)} \sum_{m=1}^{M} \frac{I(n_m > 1)}{n_m - 1} \sum_{i \in S \cap U_m} (Z_i - \tilde{Z}_m)^2. \tag{45}$$

Note also that

$$\sum_{m=1}^{M} \sum_{j \in U_m} \frac{1}{\pi_j} \mathbb{E}(Z_j - I_j \underline{x}_m^{\mathrm{T}} \underline{\beta})^2 = N(v^2 + \sigma^2).$$

The above identity via the method of moments can be used to derive an estimating equation for estimator $\check{v}^2$ of v^2

$$\check{v}^2 = f(\underline{Z}, \underline{I}, \underline{\check{\beta}}) - \check{\sigma}^2, \tag{46}$$

where $\underline{\check{\beta}}$ is an estimator of $\underline{\beta}$ and

$$f(\underline{Z}, \underline{I}, \underline{\check{\beta}}) = \frac{1}{N}\sum_{m=1}^{M}\sum_{j\in U_m}\frac{(Z_j - I_j \underline{x}_m^{\mathrm{T}}\underline{\check{\beta}})^2}{\pi_j}.$$

Note that $\mathbb{K}$, as defined in Equation (2) is a function of v^2, σ^2 and $\underline{\beta}$. Therefore, the above equation can be combined with Equation (13) into an iterative procedure as follows. We fix some starting values $\underline{\beta}_0$ and v_0^2. For any $r \geq 0$ knowing $\underline{\beta}_r$ and v_r^2, we calculate the value of $\underline{\beta}_{r+1}$ as

$$\underline{\beta}_{r+1} = (\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\mathbb{K}^{-1}(v_r^2, \check{\sigma}^2, \underline{\beta}_r)\tilde{\pi}\mathbb{X})^{-1}\mathbb{X}^{\mathrm{T}}\tilde{\pi}^{\mathrm{T}}\mathbb{K}^{-1}(v_r^2, \check{\sigma}^2, \underline{\beta}_r)\underline{Z}. \tag{47}$$

Consequently, we calculate v_{r+1}^2 according to Equation (46) as follows:

$$v_{r+1}^2 = f(\underline{Z}, \underline{I}, \underline{\beta}_{r+1}) - \check{\sigma}^2. \tag{48}$$

Though we started with a formula for an illegal pseudo-BLUE $\underline{\tilde{\beta}}$, the iterative algorithm, provided it converges, may be regarded as a valid estimation procedure. Estimators $\check{\sigma}^2$ and $\underline{\check{\beta}}$, $\check{v}^2$ obtained in such a way can be used in empirical versions of pseudo-BLUPs for θ_D and $\bar{Y}_D$ (see Equations (16) and (19))

$$\check{\theta}_D = (\underline{x}_D^{\mathrm{T}} - \check{v}^2\underline{\pi}_D^{\mathrm{T}}\check{\mathbb{K}}^{-1}\tilde{\pi}\mathbb{X})\underline{\check{\beta}} + \check{v}^2\underline{\pi}_D^{\mathrm{T}}\check{\mathbb{K}}^{-1}\underline{Z}$$

and

$$\check{Y}_D = \left(\underline{x}_D^{\mathrm{T}} - \left(\check{v}^2 + \frac{\check{\sigma}^2}{N_D}\right)\underline{\pi}_D^{\mathrm{T}}\check{\mathbb{K}}^{-1}\tilde{\pi}\mathbb{X}\right)\check{\beta} + \left(\check{v}^2 + \frac{\check{\sigma}^2}{N_D}\right)\underline{\pi}_D^{\mathrm{T}}\check{\mathbb{K}}^{-1}\underline{Z},$$

where $\check{\mathbb{K}} = \mathbb{K}(\check{v}^2, \check{\sigma}^2, \underline{\check{\beta}})$. In all simulation experiments we performed, the above numerical procedure seemed to converge fast. Unfortunately, as in many other EBLUE and EBLUPs procedures, a rigorous proof of convergence can be very difficult.

For fixed sample size designs, the above procedure for EBLUE and EBLUPs can be applied. The main difference is that we use simpler Formula (23) for $\underline{\tilde{\beta}}$, instead of Equation (13), taking into account the fact that κ_m's depend on v^2 and σ^2.

An alternative, more standard approach in this situation with $n_m > 1, m = 1, \ldots, M$, follows from the fact that conditionally on $\underline{I}$ observations $(Z_i)_{i\in S}$ are jointly independent and identically distributed within small areas. Thus, we have the following well-known relations:

$$\frac{1}{M}\sum_{m=1}^{M}\frac{1}{n_m - 1}\sum_{i\in S_m}\mathbb{E}(Z_i - \tilde{Z}_m)^2 = \sigma^2$$

and

$$\frac{1}{M-q}\sum_{m=1}^{M}\kappa_m\mathbb{E}(\tilde{Z}_m - \underline{x}^{\mathrm{T}}\underline{\tilde{\beta}})^2 = v^2.$$

In the case of fixed sample sizes considered here, the above identities via the method of moments lead to the following classical estimator of σ^2:

$$\check{\sigma}^2 = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{n_m - 1}\sum_{i\in S\cap U_m}(Z_i - \tilde{Z}_m)^2, \tag{49}$$

which is a special case of Equation (45) and estimating equation for v^2

$$\check{v}^2 = \frac{1}{M-q}\sum_{m=1}^{M}\kappa_m(\tilde{Z}_m - \underline{x}_m^{\mathrm{T}}\underline{\check{\beta}})^2, \tag{50}$$

where $\underline{\check{\beta}}$ is an estimator of $\underline{\beta}$. The numerical iterative procedure is used to compute estimators of $\underline{\beta}$ and v^2. We fix some starting value v_0^2. For any $r \geq 0$ knowing v_r^2, we calculate the value of $\underline{\beta}_{r+1}$ as

$$\underline{\beta}_{r+1} = \left(\sum_{m=1}^{M}\kappa_m(v_r^2, \check{\sigma}^2)\underline{x}_m\underline{x}_m^{\mathrm{T}}\right)^{-1}\sum_{m=1}^{M}\kappa_m(v_r^2, \check{\sigma}^2)\underline{x}_m\tilde{Z}_m, \tag{51}$$

where $\kappa_m(v^2, \sigma^2) = (v^2 n_m)/(\sigma^2 + v^2 n_m)$, $m = 1, \ldots, M$. Then v_{r+1}^2 is computed as

$$v_{r+1}^2 = \frac{1}{M-q}\sum_{m=1}^{M}\kappa_m(v_r^2, \check{\sigma}^2)(\tilde{Z}_m - \underline{x}_m^{\mathrm{T}}\underline{\beta}_{r+1})^2, \tag{52}$$

as in Equation (50).

Example 1 Assume that population $U = \{1, \ldots, N\}$ is partitioned into H strata $(V_h)_{h=1,\ldots,H}$ and each small area is a union of several strata. Without any loss of generality, we will assume that the ordering of strata is such that we can write $U_m = \bigcup_{h=h_m}^{h_{m+1}-1} V_h$ with $h_1 = 1$ and $h_{M+1} - 1 = H$. Let $N_h = \#V_h$. We consider a stratified sampling scheme selecting in each stratum V_h a sample of fixed size n_h according to simple random sampling without replacement. Thus, the sample size in each small area U_m is fixed and equal to $n_m = \bigcup_{h=h_m}^{h_{m+1}-1} n_h$. We use the formulas for BLUPs from Corollary 1. EBLUPs are constructed from BLUPs exactly as described earlier in this section. In the definition of small areas, we mimicked the situation encountered in a survey of small enterprises conducted by the Central Statistical Office in Poland. In Table 1, it is visible that, in spite of the fact that small areas are unions of strata, some of them are really very small. We expect that in the case of small areas intersecting strata, results will be qualitatively similar.

Variance component σ^2 is estimated by Equation (45) (alternatively Equation (49)). To compute the EBLUE and the estimators of variance component v^2, we use Equation (48) with

$$f(\underline{Z}, \underline{I}, \underline{\beta}) = \frac{1}{N}\sum_{m=1}^{M}\sum_{h=h_m}^{h_{m+1}-1}\frac{N_h}{n_h}\sum_{j\in V_h}(Z_j - I_j\underline{x}_m^{\mathrm{T}}\underline{\beta})^2$$

and Equation (51). The estimators of v^2 and $\underline{\beta}$ obtained in this way are denoted by $\check{v}_{\mathrm{new}}^2$ and $\underline{\check{\beta}}_{\mathrm{new}}$. The resulting predictors of θ_m and $\bar{Y}_m$ are denoted by $\check{\theta}_m^{\mathrm{new}}$ and $\check{Y}_m^{\mathrm{new}}$.

Table 1. Sizes (N_m) and sample sizes (n_m) in small areas (m).

m	1	2	3	4	5	6	7	8	9	10	11	12	13
N_m	6	5	22	51	82	186	311	458	661	844	1039	1208	1166
n_m	3	2	4	4	4	4	4	8	12	14	19	23	20
m	14	15	16	17	18	19	20	21	22	23	24	25	
N_m	1101	879	713	515	333	197	106	69	26	10	8	4	
n_m	20	16	13	8	5	4	4	4	4	4	3	2	

A more classical approach to estimating $\underline{\beta}$ and v^2, based on Equations (50) and (51), was also considered. The respective estimators and predictors are denoted by $\check{v}^2_{\text{cl}}$, $\underline{\check{\beta}}_{\text{cl}}$, $\check{\theta}^{\text{cl}}_m$ and $\check{Y}^{\text{cl}}_m$.

Our artificial population has N=10,000 units in $H = 94$ strata and $M = 25$ small areas. Denote by H_m the number of strata within the mth small area. In our case, we have $H_1 = 3$, $H_2 = 2$, $H_3 = \cdots = H_{23} = 4$, $H_{24} = 3$ and $H_{25} = 2$. Numbers of units in small areas and allocated sample sizes (proportional allocation in strata) are given in Table 1.

The total sample size is $n = 208$.

Below we present results of three experiments, each using the structure of population described above. Two of these experiments were performed under the correctly specified model, while in the third experiment the model was deliberately misspecified.

In each of the experiments, we simulate 1000 sets of population values $(Y_i)_{i\in U}$. In each simulation, we select a sample of 208 units and compute values of estimators $\check{\sigma}^2$, $\check{v}^2$ and $\underline{\check{\beta}}$ and predictors $\check{\theta}_m$ and $\check{Y}_m$ for $m = 1, \ldots, M$. In this way, we obtain the Monte Carlo approximations of expected values $\mathbb{E}\check{\sigma}^2$, $\mathbb{E}\check{v}^2$ and $\mathbb{E}\underline{\check{\beta}}$ and of MSEs:

$$\mathbb{E}(\check{\sigma}^2 - \sigma^2)^2, \quad \mathbb{E}(\check{v}^2 - v^2)^2, \quad \mathbb{E}(\underline{\check{\beta}} - \underline{\beta})^2.$$

Analogously, we approximate

$$e(\theta_m) = \mathbb{E}(\check{\theta}_m - \theta_m)^2, \quad e(\bar{Y}_m) = \mathbb{E}(\check{Y}_m - \bar{Y}_m)^2.$$

Moreover estimators, denoted, respectively, by $\hat{e}(\theta_m)$ and $\hat{e}(\bar{Y}_m)$, of the above quantities are computed via Formulas (26) and (28) with estimated values of v^2 and σ^2. The results of our experiments reported below contain Monte Carlo approximations of $\bar{\hat{e}}(\theta_m) = \mathbb{E}\hat{e}(\theta_m)$ and $\bar{\hat{e}}(\bar{Y}_m) = \mathbb{E}\hat{e}(\bar{Y}_m)$, expectations of estimators of the MSE. For brevity, we restrict ourselves to these quantities and omit the MSE of estimators of MSE of the BLUPs when summarizing the results. All the estimators and predictors are computed in two versions: classical and new (let us note that estimators of σ^2 are the same for the two approaches, $\check{\sigma}^2_{\text{cl}} = \check{\sigma}^2_{\text{new}}$).

Apart from the MSEs, we also consider the nominal 95% confidence intervals based on the normal approximation

$$\check{\theta}_m \pm 1.96\sqrt{\hat{e}(\theta_m)}, \quad \check{Y}_m \pm 1.96\sqrt{\hat{e}(\bar{Y}_m)}.$$

Expected half-widths of the intervals are denoted by

$$C^w(\theta_m) = 1.96\mathbb{E}\sqrt{\hat{e}(\theta_m)}, \quad C^w(\bar{Y}_m) = 1.96\mathbb{E}\sqrt{\hat{e}(\bar{Y}_m)}.$$

We approximate the actual coverage probabilities

$$C^{\text{pr}}(\theta_m) = \Pr(\check{\theta}_m - 1.96\sqrt{\hat{e}(\theta_m)} \le \theta_m \le \check{\theta}_m + 1.96\sqrt{\hat{e}(\theta_m)}),$$
$$C^{\text{pr}}(\bar{Y}_m) = \Pr(\check{Y}_m - \sqrt{\hat{e}(\bar{Y}_m)} \le \bar{Y}_m \le \check{Y}_m + \sqrt{\hat{e}(\bar{Y}_m)})$$

by also using the Monte Carlo method.

7.1. *Experiment 1*

We generate populations according to the formula $Y_i = \beta + u_m + \varepsilon_i$ for any $i \in U_m$ with $\beta = 50$, $v = 2$ and $\sigma = 1$. This is a special case of the general model defined in Section 3, with $q = 1$

Table 2. Comparison of simulated and estimated MSE of classical and new EBLUPs in small areas, Experiment 1.

m	$e_{\mathrm{cl}}(\theta_m)$	$\bar{\hat{e}}_{\mathrm{cl}}(\theta_m)$	$e_{\mathrm{new}}(\theta_m)$	$\bar{\hat{e}}_{\mathrm{new}}(\theta_m)$	$e_{\mathrm{cl}}(\bar{Y}_m)$	$\bar{\hat{e}}_{\mathrm{cl}}(\bar{Y}_m)$	$e_{\mathrm{new}}(\bar{Y}_m)$	$\bar{\hat{e}}_{\mathrm{new}}(\bar{Y}_m)$
1	0.317	0.304	0.324	0.300	0.166	0.159	0.169	0.158
2	0.427	0.438	0.441	0.431	0.279	0.277	0.284	0.274
3	0.225	0.233	0.228	0.230	0.189	0.193	0.191	0.191
4	0.252	0.233	0.254	0.230	0.235	0.216	0.237	0.214
5	0.237	0.233	0.241	0.230	0.226	0.222	0.230	0.220
6	0.245	0.233	0.250	0.230	0.238	0.228	0.242	0.226
7	0.253	0.233	0.256	0.230	0.253	0.230	0.256	0.228
8	0.120	0.120	0.121	0.119	0.119	0.118	0.120	0.117
9	0.080	0.081	0.080	0.081	0.079	0.080	0.079	0.079
10	0.075	0.070	0.075	0.069	0.074	0.068	0.074	0.068
11	0.052	0.052	0.052	0.051	0.051	0.051	0.051	0.050
12	0.045	0.043	0.045	0.043	0.044	0.042	0.044	0.042
13	0.050	0.049	0.05	0.049	0.049	0.048	0.049	0.048
14	0.054	0.049	0.054	0.049	0.053	0.048	0.053	0.048
15	0.063	0.061	0.063	0.061	0.061	0.06	0.061	0.060
16	0.070	0.075	0.071	0.075	0.070	0.074	0.071	0.073
17	0.114	0.120	0.115	0.119	0.113	0.118	0.113	0.118
18	0.200	0.188	0.203	0.187	0.194	0.186	0.197	0.184
19	0.229	0.233	0.231	0.230	0.223	0.228	0.225	0.226
20	0.255	0.233	0.258	0.230	0.246	0.224	0.248	0.222
21	0.239	0.233	0.240	0.230	0.226	0.220	0.226	0.218
22	0.235	0.233	0.237	0.230	0.199	0.199	0.201	0.197
23	0.251	0.233	0.255	0.230	0.151	0.143	0.153	0.142
24	0.307	0.304	0.312	0.300	0.210	0.196	0.213	0.195
25	0.465	0.438	0.472	0.431	0.245	0.234	0.248	0.232

and the auxiliary variable being constant, $x_m \equiv 1$. Note that $\theta_m = \beta + u_m$. Then, we use the same model to compute estimators and predictors.

For $\check{\sigma}^2$ both methods give the same result: $\mathbb{E}(\check{\sigma}^2 - \sigma^2)^2 \approx 0.0228$. For $\check{v}^2$ the classical method gives $\mathbb{E}(\check{v}_{\mathrm{cl}}^2 - v^2)^2 = 1.506$, while the new one is slightly worse $\mathbb{E}(\check{v}_{\mathrm{new}}^2 - v^2)^2 = 2.499$. The results for the EBLUE of β are the following: $\mathbb{E}(\check{\beta}_{\mathrm{cl}} - \beta)^2 = 0.164$ and $\mathbb{E}(\check{\beta}_{\mathrm{new}} - \beta)^2 = 0.164$. The results concerning MSEs and confidence intervals, which are specific for small areas, are gathered in Tables 2 and 3. Table 2 reports the MSE of BLUPs and average values of estimators of the MSE, while Table 3 gives the coverage rate and average half-widths of confidence intervals.

7.2. *Experiment 2*

We generate populations according to the formula $Y_i = \underline{x}_m^{\mathrm{T}}\underline{\beta} + u_m + \varepsilon_i$ for any $i \in U_m$ with $q = 2$, $\underline{\beta} = (50, 5)^{\mathrm{T}}$, $v = 2$ and $\sigma = 1$. Vectors of auxiliary variables are of the form $\underline{x}_m = (1, x_m)^{\mathrm{T}}$, with x_m generated from the standard normal distribution. The collection of x_m's is created once and kept fixed in all 1000 repeated simulations of Y_i's. Note that $\theta_m = 50 + 5x_m + u_m$. We use the correctly specified model to compute estimators and predictors.

The average value of $\check{\sigma}^2$ is 0.992, the same for both methods (the true value being $\sigma^2 = 1$). For $\check{v}_{\mathrm{cl}}^2$ the average value is 3.998, while the average value of $\check{v}_{\mathrm{new}}^2$ equals 3.785 (the true value being $v^2 = 4$).

Results for the EBLUE of $\underline{\beta}$, i.e. the Monte Carlo approximations of $\mathbb{E}\check{\underline{\beta}}$, are the following: $(49.993, 4.995)^{\mathrm{T}}$ for $\check{\underline{\beta}}_{\mathrm{cl}}$ and $(49.993, 4.995)^{\mathrm{T}}$ for $\check{\underline{\beta}}_{\mathrm{new}}$ (the true value being $\underline{\beta} = (50, 5)^{\mathrm{T}}$). Area-specific results are given in Tables 4 and 5.

Table 3. Simulated coverage rates and average half-widths of confidence intervals, Experiment 1.

m	$C^{pr}_{cl}(\theta_m)$	$C^{w}_{cl}(\theta_m)$	$C^{pr}_{new}(\theta_m)$	$C^{w}_{new}(\theta_m)$	$C^{pr}_{cl}(\bar{Y}_m)$	$C^{w}_{cl}(\bar{Y}_m)$	$C^{pr}_{new}(\bar{Y}_m)$	$C^{w}_{new}(\bar{Y}_m)$
1	0.94	1.078	0.939	1.071	0.944	0.779	0.943	0.777
2	0.948	1.295	0.944	1.284	0.945	1.029	0.94	1.024
3	0.951	0.943	0.95	0.939	0.953	0.858	0.949	0.855
4	0.935	0.943	0.935	0.939	0.931	0.908	0.934	0.904
5	0.942	0.943	0.939	0.939	0.946	0.921	0.937	0.917
6	0.939	0.943	0.938	0.939	0.949	0.933	0.946	0.929
7	0.946	0.943	0.943	0.939	0.946	0.937	0.943	0.933
8	0.95	0.677	0.948	0.676	0.949	0.672	0.946	0.67
9	0.95	0.556	0.951	0.555	0.947	0.551	0.947	0.55
10	0.939	0.516	0.938	0.515	0.943	0.511	0.942	0.511
11	0.949	0.444	0.946	0.443	0.947	0.44	0.944	0.439
12	0.943	0.404	0.943	0.403	0.946	0.4	0.946	0.399
13	0.941	0.433	0.943	0.432	0.944	0.429	0.946	0.428
14	0.939	0.433	0.94	0.432	0.935	0.429	0.935	0.428
15	0.954	0.483	0.955	0.482	0.947	0.478	0.946	0.478
16	0.955	0.535	0.954	0.534	0.956	0.53	0.955	0.529
17	0.948	0.677	0.948	0.676	0.948	0.672	0.949	0.671
18	0.944	0.849	0.941	0.845	0.943	0.843	0.941	0.839
19	0.939	0.943	0.94	0.939	0.944	0.934	0.946	0.93
20	0.941	0.943	0.936	0.939	0.947	0.926	0.942	0.922
21	0.949	0.943	0.948	0.939	0.947	0.917	0.945	0.913
22	0.95	0.943	0.947	0.939	0.943	0.872	0.945	0.868
23	0.938	0.943	0.933	0.939	0.942	0.74	0.938	0.738
24	0.95	1.078	0.947	1.071	0.947	0.866	0.944	0.863
25	0.936	1.295	0.935	1.284	0.938	0.945	0.939	0.941

Table 4. Comparison of simulated and estimated MSE of classical and new EBLUPs in small areas, Experiment 2.

m	$e_{cl}(\theta_m)$	$\bar{\hat{e}}_{cl}(\theta_m)$	$e_{new}(\theta_m)$	$\bar{\hat{e}}_{new}(\theta_m)$	$e_{cl}(\bar{Y}_m)$	$\bar{\hat{e}}_{cl}(\bar{Y}_m)$	$e_{new}(\bar{Y}_m)$	$\bar{\hat{e}}_{new}(\bar{Y}_m)$
1	0.255	0.254	0.278	0.242	0.137	0.147	0.141	0.144
2	0.345	0.333	0.368	0.313	0.246	0.24	0.253	0.233
3	0.201	0.199	0.21	0.189	0.174	0.17	0.18	0.164
4	0.212	0.2	0.224	0.192	0.202	0.188	0.213	0.181
5	0.205	0.198	0.217	0.189	0.192	0.191	0.204	0.182
6	0.204	0.199	0.221	0.19	0.198	0.196	0.215	0.188
7	0.212	0.207	0.224	0.199	0.209	0.205	0.221	0.197
8	0.108	0.11	0.114	0.107	0.105	0.109	0.111	0.105
9	0.083	0.076	0.088	0.074	0.082	0.075	0.086	0.073
10	0.067	0.066	0.069	0.065	0.066	0.066	0.067	0.064
11	0.051	0.05	0.054	0.049	0.051	0.049	0.053	0.048
12	0.043	0.042	0.044	0.041	0.042	0.041	0.043	0.04
13	0.046	0.048	0.048	0.047	0.045	0.047	0.047	0.046
14	0.045	0.048	0.046	0.047	0.044	0.047	0.045	0.046
15	0.06	0.059	0.061	0.057	0.058	0.058	0.06	0.056
16	0.07	0.071	0.073	0.069	0.07	0.07	0.073	0.068
17	0.112	0.111	0.118	0.107	0.11	0.109	0.117	0.106
18	0.17	0.165	0.18	0.158	0.169	0.163	0.178	0.156
19	0.21	0.199	0.219	0.19	0.208	0.196	0.216	0.187
20	0.218	0.212	0.23	0.205	0.21	0.206	0.222	0.199
21	0.214	0.198	0.226	0.188	0.206	0.189	0.216	0.181
22	0.203	0.198	0.225	0.188	0.18	0.174	0.194	0.167
23	0.209	0.202	0.222	0.193	0.138	0.133	0.143	0.13
24	0.256	0.247	0.276	0.233	0.177	0.175	0.187	0.17
25	0.381	0.342	0.418	0.322	0.227	0.211	0.238	0.206

Table 5. Simulated coverage rates and average half-widths of confidence intervals, Experiment 2.

m	$C^{\text{pr}}_{\text{cl}}(\theta_m)$	$C^{w}_{\text{cl}}(\theta_m)$	$C^{\text{pr}}_{\text{new}}(\theta_m)$	$C^{w}_{\text{new}}(\theta_m)$	$C^{\text{pr}}_{\text{cl}}(\bar{Y}_m)$	$C^{w}_{\text{cl}}(\bar{Y}_m)$	$C^{\text{pr}}_{\text{new}}(\bar{Y}_m)$	$C^{w}_{\text{new}}(\bar{Y}_m)$
1	0.95	0.986	0.93	0.96	0.957	0.75	0.949	0.742
2	0.944	1.128	0.917	1.089	0.945	0.959	0.932	0.944
3	0.945	0.871	0.932	0.849	0.951	0.807	0.931	0.791
4	0.938	0.876	0.927	0.855	0.931	0.849	0.92	0.831
5	0.945	0.87	0.918	0.848	0.945	0.854	0.929	0.834
6	0.943	0.873	0.92	0.851	0.942	0.866	0.924	0.845
7	0.943	0.889	0.932	0.871	0.94	0.884	0.923	0.867
8	0.949	0.65	0.937	0.638	0.947	0.645	0.938	0.634
9	0.94	0.541	0.933	0.533	0.939	0.536	0.928	0.529
10	0.947	0.504	0.942	0.498	0.949	0.5	0.942	0.494
11	0.943	0.437	0.937	0.432	0.944	0.433	0.943	0.428
12	0.945	0.399	0.938	0.396	0.946	0.396	0.943	0.393
13	0.953	0.426	0.947	0.422	0.953	0.423	0.944	0.419
14	0.962	0.426	0.955	0.422	0.96	0.422	0.954	0.418
15	0.945	0.474	0.941	0.468	0.95	0.469	0.945	0.464
16	0.947	0.522	0.943	0.515	0.944	0.517	0.938	0.511
17	0.946	0.65	0.935	0.639	0.95	0.646	0.941	0.635
18	0.938	0.795	0.923	0.777	0.943	0.79	0.925	0.772
19	0.939	0.872	0.929	0.851	0.941	0.866	0.928	0.845
20	0.946	0.901	0.937	0.885	0.949	0.887	0.945	0.872
21	0.935	0.869	0.919	0.847	0.936	0.85	0.913	0.83
22	0.948	0.869	0.925	0.846	0.948	0.816	0.929	0.799
23	0.942	0.878	0.93	0.858	0.935	0.712	0.927	0.704
24	0.946	0.971	0.919	0.942	0.949	0.817	0.934	0.805
25	0.937	1.142	0.907	1.105	0.942	0.898	0.929	0.887

7.3. *Experiment 3*

This experiment is designed to investigate the behaviour of estimators and predictors under a misspecified model and to assess the predictive power of auxiliary variables. The 'true' model used to generate populations is exactly the same as in Experiment 2, that is, $Y_i = \underline{x}_m^{\mathrm{T}}\underline{\beta} + u_m + \varepsilon_i$ for any $i \in U_m$ with $q = 2$, $\underline{\beta} = (50, 5)^{\mathrm{T}}$, $v = 2$ and $\sigma = 1$. Thus, $\theta_m = 50 + 5x_m + u_m$ and $Y_i = \theta_m + \varepsilon_i$ for any $i \in U_m$.

In contrast to Experiment 2, when computing estimators/predictors, we assume that a statistician uses an incorrect model without auxiliary variables, that is, she works under a 'wrong' assumption that $Y_i = \beta + u_m + \varepsilon_i$ with scalar intercept β, as in Experiment 1. There are several reasons to consider such a situation. First, in reality the 'true' model never exists and the statistician has always to make inferences based on simplified assumptions. Second, auxiliary variables can be either inaccessible or the cost of collecting them can be significant. Then a statistician may be forced to use a model which does not require any knowledge of these variables. It is interesting to examine how misspecification of the model or inaccessibility of x_i's affects the accuracy of the predictors. Results reported below should be interpreted with some caution, because when a model is misspecified then speaking of 'true values' of some estimated parameters may be misleading.

The average value of $\check{\sigma}^2$ is 1, the same for both methods, and practically equal to the 'true' value $\sigma^2 = 1$ used to generate data. For $\check{v}^2_{\text{cl}}$ the average value is 30.251, the average value of $\check{v}^2_{\text{new}}$ equals 16.587 (the 'true' value used to generate data being $v^2 = 4$).

Results for the EBLUE of the intercept β, i.e. Monte Carlo approximations of $\mathbb{E}\check{\beta}$ for both methods are very close to each other: 49.979 for $\check{\beta}_{\text{cl}}$ and for $\check{\beta}_{\text{new}}$ (it does not make sense to speak of the true value of β because to generate data we used the two-dimensional vector $\underline{\beta} = (50, 5)^{\mathrm{T}}$ and x_i's; it is, however, true that simulated Y_i's satisfy the relation $\mathbb{E}Y_i = 50 + 5x_i$ and the average value of x_i's is close to 0). Area-specific results are given in Tables 6 and 7.

Table 6. Comparison of simulated and estimated MSE of classical and new EBLUPs in small areas, Experiment 3.

m	$e_{\text{cl}}(\theta_m)$	$\bar{\hat{e}}_{\text{cl}}(\theta_m)$	$e_{\text{new}}(\theta_m)$	$\bar{\hat{e}}_{\text{new}}(\theta_m)$	$e_{\text{cl}}(\bar{Y}_m)$	$\bar{\hat{e}}_{\text{cl}}(\bar{Y}_m)$	$e_{\text{new}}(\bar{Y}_m)$	$\bar{\hat{e}}_{\text{new}}(\bar{Y}_m)$
1	0.35	0.33	0.364	0.327	0.173	0.166	0.176	0.165
2	0.492	0.492	0.494	0.485	0.28	0.297	0.281	0.295
3	0.26	0.248	0.26	0.246	0.215	0.203	0.215	0.202
4	0.258	0.248	0.259	0.246	0.246	0.229	0.247	0.227
5	0.243	0.248	0.241	0.246	0.237	0.236	0.235	0.234
6	0.231	0.248	0.231	0.246	0.225	0.243	0.226	0.241
7	0.248	0.248	0.263	0.246	0.245	0.245	0.26	0.243
8	0.125	0.124	0.124	0.124	0.123	0.122	0.122	0.122
9	0.079	0.083	0.079	0.083	0.078	0.082	0.078	0.081
10	0.071	0.071	0.071	0.071	0.07	0.07	0.07	0.07
11	0.053	0.052	0.053	0.052	0.052	0.052	0.052	0.052
12	0.043	0.043	0.043	0.043	0.042	0.043	0.042	0.042
13	0.048	0.05	0.048	0.05	0.047	0.049	0.047	0.049
14	0.053	0.05	0.053	0.05	0.052	0.049	0.052	0.049
15	0.065	0.062	0.065	0.062	0.064	0.061	0.064	0.061
16	0.078	0.077	0.078	0.077	0.077	0.075	0.077	0.075
17	0.131	0.124	0.131	0.124	0.128	0.123	0.128	0.122
18	0.203	0.199	0.202	0.198	0.2	0.196	0.2	0.195
19	0.234	0.248	0.235	0.246	0.228	0.243	0.228	0.241
20	0.266	0.248	0.293	0.246	0.252	0.239	0.277	0.237
21	0.243	0.248	0.24	0.246	0.235	0.234	0.233	0.232
22	0.253	0.248	0.249	0.246	0.217	0.21	0.215	0.209
23	0.254	0.248	0.26	0.246	0.151	0.149	0.153	0.149
24	0.338	0.33	0.333	0.327	0.226	0.207	0.224	0.206
25	0.519	0.492	0.536	0.485	0.254	0.248	0.259	0.246

Table 7. Simulated coverage rates and average half-widths of confidence intervals, Experiment 3.

m	$C^{\text{pr}}_{\text{cl}}(\theta_m)$	$C^{w}_{\text{cl}}(\theta_m)$	$C^{\text{pr}}_{\text{new}}(\theta_m)$	$C^{w}_{\text{new}}(\theta_m)$	$C^{\text{pr}}_{\text{cl}}(\bar{Y}_m)$	$C^{w}_{\text{cl}}(\bar{Y}_m)$	$C^{\text{pr}}_{\text{new}}(\bar{Y}_m)$	$C^{w}_{\text{new}}(\bar{Y}_m)$
1	0.94	1.122	0.931	1.117	0.944	0.796	0.945	0.794
2	0.953	1.371	0.948	1.362	0.961	1.065	0.96	1.061
3	0.941	0.973	0.938	0.97	0.929	0.881	0.932	0.878
4	0.933	0.973	0.935	0.97	0.931	0.934	0.933	0.932
5	0.958	0.973	0.958	0.97	0.943	0.949	0.944	0.946
6	0.954	0.973	0.954	0.97	0.957	0.963	0.959	0.959
7	0.951	0.973	0.944	0.97	0.949	0.967	0.941	0.964
8	0.945	0.689	0.944	0.688	0.948	0.683	0.948	0.682
9	0.955	0.563	0.956	0.563	0.956	0.558	0.956	0.558
10	0.952	0.522	0.952	0.521	0.948	0.517	0.949	0.517
11	0.944	0.448	0.945	0.448	0.943	0.444	0.942	0.443
12	0.948	0.407	0.947	0.407	0.945	0.403	0.945	0.403
13	0.956	0.437	0.954	0.436	0.954	0.433	0.953	0.433
14	0.94	0.437	0.94	0.436	0.933	0.433	0.937	0.432
15	0.932	0.488	0.93	0.488	0.932	0.484	0.937	0.483
16	0.939	0.541	0.936	0.541	0.938	0.536	0.937	0.536
17	0.94	0.689	0.94	0.688	0.943	0.684	0.941	0.683
18	0.957	0.871	0.953	0.869	0.953	0.865	0.953	0.862
19	0.955	0.973	0.951	0.97	0.963	0.963	0.957	0.96
20	0.936	0.973	0.925	0.97	0.941	0.955	0.926	0.952
21	0.948	0.973	0.949	0.97	0.939	0.945	0.939	0.942
22	0.944	0.973	0.943	0.97	0.935	0.896	0.937	0.893
23	0.949	0.973	0.945	0.97	0.946	0.755	0.948	0.753
24	0.938	1.122	0.942	1.117	0.954	0.889	0.954	0.886
25	0.943	1.371	0.94	1.362	0.935	0.973	0.927	0.97

7.4. *Conclusions*

The primary goal of our simulation study is to compare accuracy of new predictors proposed in this paper, say $\check{\theta}_m^{\text{new}}$ and $\check{Y}_m^{\text{new}}$, with the classical ones, $\check{\theta}_m^{\text{cl}}$ and $\check{Y}_m^{\text{cl}}$. Note that we restrict experiments to the case when sample sizes in small areas are non-random and, consequently, BLUPs exist, they are given by formulas in Corollary 1 and they coincide with classical ones. Therefore, the only difference between $\check{\theta}_m^{\text{new}}$ and $\check{\theta}_m^{cl}$ (respectively, $\check{Y}_m^{\text{new}}$ and $\check{Y}_m^{\text{cl}}$) is in that two different methods are used to estimate v^2 and thus to construct EBLUPs. New estimator $\check{v}_{\text{new}}^2$ turns out to be slightly worse than $\check{v}_{\text{cl}}^2$ in our experiments (compare the MSE of two estimators in Experiments 1 and 2). However, inspection of Tables 2, 4 and 6 reveals that the accuracy of new EBLUPs is very similar to that of classical EBLUPs. It seems that a different method of estimating the variance component has a negligible effect on the accuracy of final predictors. This is in full agreement with the results of Kackar and Harville [18] (see also [2, Chapter 6]). Similar conclusions concern confidence intervals. Two methods (one based on $\check{\theta}_m^{\text{new}}$ or $\check{Y}_m^{\text{new}}$ combined with $\check{v}_{\text{new}}^2$, the other on $\check{\theta}_m^{\text{cl}}$ or $\check{Y}_m^{\text{cl}}$ combined with $\check{v}_{\text{cl}}^2$, have comparable properties, that is, the similar actual coverage rate and widths of intervals. This is visible in Tables 3, 5 and 7. The actual coverage rate is slightly but systematically lower than the nominal rate 0.95 for both methods. The normal approximation, only heuristically justified in the model under consideration, seems to work remarkably well.

Comparison of results of Experiment 2 with those of Experiment 3 allows us to draw some conclusions which are interesting from the practical viewpoint, but not directly pertaining to the approach proposed in our paper. In fact, in both experiments, the difference between new and classical predictors is small, as we already mentioned. What is really striking, however, is the relatively small difference *between accuracy of predictors in Experiment* 2 *and those in Experiment* 3. Entries of Table 4 corresponding to middle rows (relatively big 'small areas') are *not* significantly different from those of Table 6! It is true that the MSE of predictors in 'really small' areas, in top and bottom rows in Table 4, are smaller than those in Table 6, but the difference is much less than we expected. Let us recall that in Experiment 2, we use the correctly specified model and assume that values of auxiliary variables are known. In Experiment 3, the model is misspecified and auxiliary variables *ignored.* Comparison of Table 5 with Table 7 leads to similar, rather surprising conclusions. Confidence intervals in Experiment 2 are significantly narrower than those in Experiment 3 only for very small areas, while quite similar for bigger ones. Similar results (sometimes in an even more pronounced form) were observed also in other experiments, which will not be reported because of space limitations.

The basic small area model considered here is very robust against misspecification of some of its elements. We offer the following heuristic explanation of this phenomenon. If the data obey the model $Y_i = \beta + \beta_1 x_m + u_m + \varepsilon_i$ while the analysis is performed as if they were of the form $Y_i = \beta + u_m + \varepsilon_i$, then the extra variability due to the missing term $\beta_1 x_m$ is absorbed into the estimator of v^2, the variance of u_m. In Experiment 3, the average value of $\check{v}_{\text{cl}}^2$ is 30.251, very far from $v^2 = 4$ (the value used when generating u_m's) but remarkably close to $v^2 + \beta_1^2 \sum(x_m - \bar{x})^2/M$ (recall that artificial x_m's are generated from the standard normal distribution and $\beta_1^2 = 25$). As a result, estimated/predicted small area effects u_m seem to 'take over the role' played by omitted terms $\beta_1 x_m$. Of course, predictors which make no use of x_m's should be worse than those which take x_m's into account, and indeed in our experiment they are worse. What is not so obvious is that they are *not very much* worse, with differences visible only for very small domains. Differences between predictors $\check{Y}_m$ are even lesser than between $\check{\theta}_m$, due to the extra variability contributed by unobserved units.

Summing up, EBLUPs in our experiments perform very well even if the model is misspecified. The predictive power of auxiliary variables turns out to be not as decisive as we initially expected. Of course, these conclusions should be applied in practice with caution, because our simple experiments do not reflect all the complexity of a real application. However, we think that the results reported above point out to a problem which deserves further investigation.

8. Bernoulli sampling

In this section, we discuss the most elementary example of a sampling plan with a random sample size, which is the Bernoulli sampling. In this case, general formulas have to be applied for BLUE and BLUPs. However, relatively simple analytical expressions for them can be obtained, at least in the case of univariate auxiliary variables.

For the Bernoulli sampling $b(\pi)$, each element is taken to the sample independently with a fixed probability $\pi \in (0, 1)$, i.e. $\underline{\pi} = \pi \underline{1}$, $\Pi = \pi \mathbb{I}_N$, $\underline{\pi}_m = \pi \underline{e}_m$, $m = 1, \ldots, M$, $\underline{\tilde{\pi}} \underline{\tilde{\pi}}^{\mathrm{T}} = \pi^2 \mathrm{Diag}(\underline{e}_m \underline{e}_m^{\mathrm{T}})$ and

$$\mathbb{P} - \underline{\pi}\underline{\pi}^{\mathrm{T}} = \tilde{\mathbb{P}} - \underline{\tilde{\pi}}\underline{\tilde{\pi}}^{\mathrm{T}} = \pi(1-\pi)\mathbb{I}_N.$$

Therefore, in the model with a univariate auxiliary variable, considered in Section 5, we get

$$\mathbb{K} = \pi \mathrm{Diag}\{[\sigma^2 + (\beta^2 x_m^2 + v^2)(1-\pi)]\mathbb{I}_{N_m} + \pi v^2 \underline{e}_m \underline{e}_m^{\mathrm{T}}\}.$$

Consequently, using the block inversion, we get

$$\mathbb{K}^{-1} = \frac{1}{\pi} \mathrm{Diag}\left[\frac{1}{\sigma^2 + (\beta^2 x_m^2 + v^2)(1-\pi)}\left(\mathbb{I}_{N_m} - \frac{\pi v^2 \underline{e}_m \underline{e}_m^{\mathrm{T}}}{\sigma^2 + (\beta^2 x_m^2 + v^2)(1-\pi) + v^2 \pi N_m}\right)\right].$$

Thus

$$\mathbb{K}^{-1}\tilde{\pi}\underline{x} = \pi \sum_{m=1}^{M} x_m \mathbb{K}^{-1} \underline{e}_m = \sum_{m=1}^{M} \frac{x_m \underline{e}_m}{\sigma^2 + (\beta^2 x_m^2 + v^2)(1-\pi) + v^2 \pi N_m}.$$

Furthermore,

$$\underline{x}^{\mathrm{T}} \tilde{\pi}^{\mathrm{T}} \mathbb{K}^{-1} \tilde{\pi} \underline{x} = \pi \sum_{m=1}^{M} \frac{x_m^2 N_m}{\sigma^2 + (\beta^2 x_m^2 + v^2)(1-\pi) + v^2 \pi N_m}.$$

Moreover,

$$\underline{x}^{\mathrm{T}} \tilde{\pi}^{\mathrm{T}} \mathbb{K}^{-1} \underline{Z} = \sum_{m=1}^{M} \frac{x_m N_m \tilde{Z}_m}{\sigma^2 + (\beta^2 x_m^2 + v^2)(1-\pi) + v^2 \pi N_m},$$

where $\tilde{Z}_m = \sum_{i \in U_m} Z_i / (\pi N_m)$, $m = 1, \ldots, M$. Consequently, by Equation (33), the BLUE of β has the form

$$\hat{\beta} = \frac{\sum_{m=1}^{M} \frac{x_m N_m \tilde{Z}_m}{\tau^2 + \left(\frac{1+\tau^2}{\gamma^2} x_m^2 + 1\right)(1-\pi) + \pi N_m}}{\sum_{m=1}^{M} \frac{x_m^2 N_m}{\tau^2 + \left(\frac{1+\tau^2}{\gamma^2} x_m^2 + 1\right)(1-\pi) + \pi N_m}} \tag{53}$$

and its MSE is

$$\mathbb{E}(\hat{\beta} - \beta)^2 = \frac{1}{\pi \sum_{m=1}^{M} (N_m / (\sigma^2 + (\beta^2 + v^2)(1-\pi) + v^2 \pi N_m))}.$$

To find BLUPs $\hat{\theta}_D$ and $\hat{Y}_D$, we need also the formula

$$\underline{x}^{\mathrm{T}} \tilde{\pi}^{\mathrm{T}} \bar{\mathbb{K}}^{-1} \underline{\pi}_D = \frac{\pi v^2 x_D N_D}{\sigma^2 + (\beta^2 x_D^2 + v^2)(1-\pi) + v^2 \pi N_D}.$$

Therefore, in view of Equations (35) and (53), we obtain the BLUP of θ_D as follows

$$\hat{\theta}_D = \frac{x_D\left[\tau^2 + \left(\frac{1+\tau^2}{\gamma^2}x_D^2 + 1\right)(1-\pi)\right]}{\tau^2 + \left(\frac{1+\tau^2}{\gamma^2}x_D^2 + 1\right)(1-\pi) + \pi N_D} \times \frac{\sum_{m=1}^{M} \frac{x_m N_m}{\tau^2 + \left(\frac{1+\tau^2}{\gamma^2}x_m^2+1\right)(1-\pi)+\pi N_M}\tilde{Z}_m}{\sum_{m=1}^{M} \frac{x_m^2 N_m}{\tau^2 + \left(\frac{1+\tau^2}{\gamma^2}x_m^2+1\right)(1-\pi)+\pi N_m}}$$

$$+ \frac{\pi N_D}{\tau^2 + \left(\frac{1+\tau^2}{\gamma^2}x_D^2 + 1\right)(1-\pi) + \pi N_D}\tilde{Z}_D.$$

That is

$$\hat{\theta}_D = \delta_D x_D \hat{\beta} + (1-\delta_D)\tilde{Z}_D$$

with

$$\delta_D = \frac{\tau^2 + (((1+\tau^2)/\gamma^2)x_D^2 + 1)(1-\pi)}{\tau^2 + (((1+\tau^2)/\gamma^2)x_D^2 + 1)(1-\pi) + \pi N_D}.$$

Its MSE, due to Equation (36), is

$$\mathbb{E}(\hat{\theta}_D - \theta_D)^2 = \delta_D v^2 \frac{\sum_{m=1}^{M} x_m^2 - \sum_{m=1, m\neq D}^{M} x_m^2 \delta_m}{\sum_{m=1}^{M} x_m^2(1-\delta_m)}.$$

Finally, the BLUP of $\bar{Y}_D$ has the form

$$\hat{Y}_D = \frac{(\tau^2 + \frac{1+\tau^2}{\gamma^2}x_D^2 + 1)(1-\pi)}{\tau^2 + \left(\frac{1+\tau^2}{\gamma^2}x_D^2 + 1\right)(1-\pi) + v^2\pi N_D} \times \frac{\sum_{m=1}^{M} \frac{x_m N_m}{\tau^2 + \left(\frac{1+\tau^2}{\gamma^2}x_D^2+1\right)(1-\pi)+\pi N_M}\tilde{Z}_m}{\sum_{m=1}^{M} \frac{x_m^2 N_m}{\tau^2 + \left(\frac{1+\tau^2}{\gamma^2}x_D^2+1\right)(1-\pi)+\pi N_m}}$$

$$+ \frac{\pi(\tau^2 + N_D)}{\tau^2 + \left(\frac{1+\tau^2}{\gamma^2}x_D^2 + 1\right)(1-\pi) + \pi N_D}\tilde{Z}_D.$$

Thus

$$\hat{Y}_D = \alpha_D x_D \hat{\beta} + (1-\alpha_D)\tilde{Z}_D,$$

with $\alpha_D = 1 - \tau_D(1-\delta_D)$. Its MSE is

$$\mathbb{E}(\hat{Y}_D - \bar{Y}_D)^2 = \alpha_D \frac{x_D^2 + \tau_D \sum_{m\neq D} x_m^2(1-\delta_m)}{\sum_{m=1}^{M} x_m^2(1-\delta_m)}.$$

To find EBLUE $\check{\beta}$ of β and estimators of variance components, we proceed as follows. For the estimator of σ^2, we use Equation (45) with

$$\sum_{m=1}^{M} \Pr(n_m > 1) = M - \sum_{m=1}^{M}(1 + \pi N_m)(1-\pi)^{N_m - 1}.$$

We find $\check{\beta}$ and $\check{v}$ simultaneously using analogues of Equations (46) and (47), that is, we start with fixing some v_0^2 and β_0. Then knowing v_r^2 and β_r, the new value β_{r+1} is calculated via

$$\beta_{r+1} = \frac{\sum_{m=1}^{M} \frac{x_m N_m \tilde{Z}_m}{\check{\sigma}^2 + (\beta_r^2 x_m^2 + v_r^2)(1-\pi) + v_r^2 \pi N_m}}{\sum_{m=1}^{M} \frac{x_m^2 N_m}{\check{\sigma}^2 + (\beta_r^2 x_m^2 + v_r^2)(1-\pi) + v_r^2 \pi N_m}}.$$

Consequently, υ_{r+1}^2 is calculated by

$$\upsilon_{r+1}^2 = \frac{1}{\pi N}\sum_{m=1}^{M}\sum_{j\in U_m}(Z_j - I_j x_m \beta_{r+1})^2 - \breve{\sigma}^2.$$

The procedure converges rather fast to $\breve{\upsilon}^2$ and $\breve{\beta}$.

Using $\breve{\beta}$ and estimators of variance components $\breve{\sigma}^2$ and $\breve{\upsilon}^2$, we arrive at the form of the EBLUP of θ_D as follows:

$$\breve{\theta}_D = \breve{\delta}_D x_D \breve{\beta} + (1-\breve{\delta}_D)\tilde{Z}_D,$$

where

$$\breve{\delta}_D = \frac{\breve{\sigma}^2 + (\breve{\beta}^2 x_D^2 + \breve{\upsilon}^2)(1-\pi)}{\breve{\sigma}^2 + (\breve{\beta}^2 x_D^2 + \breve{\upsilon}^2)(1-\pi) + \breve{\upsilon}^2 \pi N_D}.$$

Again using $\breve{\beta}$ and estimators of variance components $\breve{\upsilon}^2$ and $\breve{\sigma}^2$, we get the EBLUP of $\bar{Y}_D$ as

$$\breve{Y}_D = \breve{\alpha}_D x_D \breve{\beta} + (1-\breve{\alpha}_D)\tilde{Z}_D,$$

where $\breve{\alpha}_D = 1 - \breve{\tau}_D(1-\breve{\delta}_D)$ with $\breve{\tau}_D = 1 + \breve{\sigma}^2/\breve{\upsilon}^2 N_D$.

Using these numerical formulas, one can derive values of predictors of θ_D and $\bar{Y}_D$. Of course, their quality depends strongly on the true value of β. In numerical experiments, we performed for $\pi = 0.05$ in a population of the size N=10,000 reasonable values of predictors were obtained for rather large β and small variances of u_m's and ε_i's. Moreover, results were in most cases worse than those obtained via the method designed for the case of fixed sample sizes, described in formulas (49)–(52). The results obtained by both methods were comparable in the case of large coefficient of variation of Y_i's.

Though these numerical results are slightly disappointing, they do not contradict our theoretical optimality results because classical predictors given by Equations (41) and (43) are not linear if the design plan is not of fixed sample sizes (e.g. in the case of the Bernoulli sampling). They are only conditionally linear, given n_m's.

Acknowledgements

This work was partially supported by Polish Ministry of Science and Higher Education Grant No. 1 H02B 02 30

References

[1] R.E. Fay and R. Herriot, *Estimates of income for small places: An application of James–Stein procedure to census data*, J. Amer. Statist. Assoc. 68 (1979), pp. 269–277.

[2] J.N.K. Rao, *Small Area Estimation*, Wiley, Hoboken, NJ, 2003.

[3] D.A. Marker, *Organization of small area estimators using a generalized linear regression framework*, J. Official Stat. 15(1) (1999), pp. 1–24.

[4] M. Ghosh and J.N.K. Rao, *Small area estimation: An appraisal*, Statist. Sci. 9 (1994), pp. 55–93.

[5] M. Ghosh, *Model dependent small area estimation – theory and practice*, in *Lecture Notes on Estimation for Population Domains and Small Areas*, R. Lehtonen and K. Djerf, eds., Statistics Finland, Helsinki, 2001, pp. 51–108.

[6] R. Chambers, *An introduction to model-based survey sampling*, in *Workshop on Survey Sampling Theory and Methodology*, D. Krapavickaite and A. Plikusas, eds., Statistics Lithuania, Vilnius, 2005, pp. 9–51.

[7] R.M. Royall, *On finite population sampling theory under certain linear regression models*, Biometrika 57(2) (1970), pp. 377–387.

[8] C. Cassel, C.-E. Särndal, and J. Wretman, *Foundations of Inference in Survey Sampling*, Wiley, New York, 1977.

[9] D.R. Bellhouse, *A review of optimal designs in survey sampling*, Can. J. Stat. 12(1) (1984), pp. 53–65.

[10] H. Chandra and R. Chambers, *Comparing EBLUP and C-EBLUP for small area estimation*, Stat. Transit. 7(3) (2005), pp. 637–648.

[11] D. Pfeffermann, *Small area estimation – new developments and directions*, Int. Statist. Rev. 70 (2002), pp. 125–143.
[12] W. Niemiro and J. Wesołowski, *Synthetic and composite estimation under a superpopulation model*, Statist. Pap. 51(3) (2010), pp. 497–509.
[13] A. Scott and T.M.F. Smith, *A note on estimating secondary characteristics in multivariate surveys*, Sankhyã, Ser. A 31 (1969), pp. 497–499.
[14] C.-E. Särndal, *Design-based methodologies for domain estimation*, in *Lecture Notes on Estimation for Population Domains and Small Areas*, R. Lehtonen and K. Djerf, eds., Statistics Finland, Helsinki, 2001, pp. 5–49.
[15] D.A. Marker, *Producing small area estimates from national surveys: Methods for minimizing use of indirect estimators*, Surv. Methodol. 27(2), (2001), pp. 183–188.
[16] M.P. Singh, J. Gambino, and H.J. Mantel, *Issues and strategies for small area data*, Surv. Methodol. 20 (1994), pp. 3–14.
[17] C.-E. Särndal, B. Swensson, and J. Wretman, *Model Assisted Survey Sampling*, Springer-Verlag, New York, 1992.
[18] R.N. Kackar and D.A. Harville, *Approximations for standard errors of estimators of fixed and random effects in mixed linear models*, J. Amer. Statist. Assoc. 79 (1984), pp. 853–862.