Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Measuring uncertainty associated with model-based small area estimators

by J.N.K. Rao, Susana Rubin-Bleuer and Victor M. Estevao

Release date: December 20, 2018

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p} preliminary
^{r} revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E} use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

An HTML version **is also available.**

Cette publication est aussi disponible en français.

Survey Methodology, December 2018
Vol. 44, No. 2, pp. 151-166
Statistics Canada, Catalogue No. 12-001-X

Measuring uncertainty associated with model-based small area estimators

J.N.K. Rao, Susana Rubin-Bleuer and Victor M. Estevao[1]

Abstract

Domains (or subpopulations) with small sample sizes are called small areas. Traditional direct estimators for small areas do not provide adequate precision because the area-specific sample sizes are small. On the other hand, demand for reliable small area statistics has greatly increased. Model-based indirect estimators of small area means or totals are currently used to address difficulties with direct estimation. These estimators are based on linking models that borrow information across areas to increase the efficiency. In particular, empirical best (EB) estimators under area level and unit level linear regression models with random small area effects have received a lot of attention in the literature. Model mean squared error (MSE) of EB estimators is often used to measure the variability of the estimators. Linearization-based estimators of model MSE as well as jackknife and bootstrap estimators are widely used. On the other hand, National Statistical Agencies are often interested in estimating the design MSE of EB estimators in line with traditional design MSE estimators associated with direct estimators for large areas with adequate sample sizes. Estimators of design MSE of EB estimators can be obtained for area level models but they tend to be unstable when the area sample size is small. Composite MSE estimators are proposed in this paper and they are obtained by taking a weighted sum of the design MSE estimator and the model MSE estimator. Properties of the MSE estimators under the area level model are studied in terms of design bias, relative root mean squared error and coverage rate of confidence intervals. The case of a unit level model is also examined under simple random sampling within each area. Results of a simulation study show that the proposed composite MSE estimators provide a good compromise in estimating the design MSE.

Key Words: Area and unit level models; Composite estimators of design mean squared error; Empirical best linear unbiased predictor; Estimating design mean squared error.

1 Introduction

Sample survey data are often used to produce estimates of domain (subpopulation) totals or means. Traditional direct estimators for domains, including calibration estimators that use known population totals of auxiliary variables, are designed to provide reliable estimators for domains with large domain-specific sample sizes. However, direct estimators do not provide adequate precision for domains with small sample sizes (called small areas). Yet the demand for reliable small area statistics has greatly increased in recent years. It is therefore necessary to resort to indirect estimators that borrow information from related areas through known auxiliary information such as censuses and administrative records, to increase the efficiency. Indirect estimators based on explicit linking models are widely used; in particular, empirical best (EB) estimators based on area level or unit level linear regression models with random area effects. A detailed account of EB estimation under those models is given by Rao and Molina (2015), Chapters 6 and 7. Section 2 presents EB estimators of small area means under basic area level and unit level models.

EB-type model based estimators are often deemed suitable by National Statistical Agencies to produce official statistics, after careful external evaluations. For example, Beaumont and Bocci (2016) compared EB and direct estimates of unemployment rate for small areas obtained from the Canadian Labour Force Survey

1. J.N.K. Rao, Distinguished Research Professor, Carleton University, Ottawa, Ontario. E-mail: jrao34@rogers.com; Susana Rubin-Bleuer, Adjunct Research Professor, Carleton University, Ottawa, Ontario; Victor M. Estevao, Senior Methodologist, Statistics Canada, R.H. Coats Bldg, 25th Floor, 100 Tunney's Pasture, Ottawa, Ontario, K1A 0T6. E-mail: victor.estevao@canada.ca.

(LFS) to "gold standard" estimates obtained from the much larger National Household Survey (comparable to long form census) and found that the relative error of EB estimates is much smaller than the corresponding direct estimates. The authors used a basic area level linear regression model with random area effects to produce EB estimates. External evaluations were first used in the pioneering paper by Fay and Herriot (1979) under a basic area level model to produce estimates of mean income for small places in the United States.

Model mean squared error (MSE) of the EB estimators is often used to measure the variability of the estimators. In particular, linearization-based estimators of model MSE as well as jackknife and bootstrap estimators are widely used. Section 3 gives a brief account of model-based MSE estimation, including estimators based on unconditional and conditional frameworks.

The literature on estimating model MSE is very impressive, but National Statistical agencies are often interested in estimating the design MSE of EB estimators in line with the traditional design MSE estimators of direct estimators for large areas with adequate sample sizes (Pfeffermann and Gilboa, 2017). Estimators of design MSE of EB estimators for the basic area level model can be obtained but they tend to be unstable when the area sample size is small. To address this problem, Section 4 proposes composite MSE estimators obtained by taking a weighted sum of the design MSE estimator and the model MSE estimator. The case of unit level models is also studied under simple random sampling within areas. Section 5 reports the results of simulation studies on the performance of the proposed composite MSE estimators in terms of design absolute relative bias (ARB), relative root mean squared error (RRMSE) and coverage of confidence intervals. Both area level and unit level models are considered in the simulation study. Finally, some conclusions are given in Section 6.

2 EB estimators

In this section, we present EB estimators of small area means or totals, denoted by θ_i, for m areas with small sample sizes. For area level models we assume that direct estimators $\hat{\theta}_i$ and associated area level covariates $\mathbf{z}_i$ are available for the m areas, where $\mathbf{z}_i$ is a $p \times 1$ vector. In the case of unit level models, we assume that unit level data $\{(y_{ij}, \mathbf{x}_{ij}), \quad j = 1, \ldots, n_i; \quad i = 1, \ldots, m\}$ are available for the sampled areas, where n_i is the sample size in area i and $\mathbf{x}_{ij}$ is a $p \times 1$ vector of covariates that can include area level covariates. We assume that the area population means $\bar{\mathbf{X}}_i$ are known.

2.1 Basic area level model

We assume that the direct estimator $\hat{\theta}_i$ is design unbiased (either exactly or approximately for large overall sample size n). For example, estimators calibrated to known overall means of auxiliary variables are approximately unbiased. We can express this assumption as a sampling model $\hat{\theta}_i = \theta_i + e_i$, where the sampling error e_i has zero mean and variance ψ_i. We further assume that the sampling variance ψ_i is known and not random. In practice, the estimators of the sampling variances are smoothed and the resulting smoothed estimator is taken as a proxy for ψ_i. Beaumont and Bocci (2016) propose a method of smoothing

the sampling variances in the context of Canadian LFS. The model linking the areas assumes that the θ_i are random, obeying the "matching" linking model $\theta_i = \mathbf{z}_i'\boldsymbol{\beta} + v_i$, where the random area effect v_i has zero mean and variance σ_v^2 and is independent of the sampling error e_i. We further assume normality of v_i and e_i.

Combining the sampling model with the linking model leads to the basic area level model

$$\hat{\theta}_i = \mathbf{z}_i'\boldsymbol{\beta} + v_i + e_i, \quad v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2), \quad e_i \overset{\text{id}}{\sim} N(0, \psi_i), \quad i = 1, \ldots, m. \tag{2.1}$$

Main advantages of model (2.1) are that it takes account of the sampling design through the sampling model on the direct estimators and that it requires only area level covariates, which are more readily available than unit level covariates.

For known model parameters $(\boldsymbol{\beta}, \sigma_v^2)$, the "best" estimator of θ_i is given by

$$\tilde{\theta}_i^B = \tilde{E}\left(\theta_i \mid \hat{\theta}_i, \boldsymbol{\beta}, \sigma_v^2\right) = \gamma_i \hat{\theta}_i + (1 - \gamma_i)\mathbf{z}_i'\boldsymbol{\beta}, \tag{2.2}$$

where $\gamma_i = \sigma_v^2/(\sigma_v^2 + \psi_i)$. The best estimator (2.2) is unbiased for θ_i in the sense that $E\left(\tilde{\theta}_i^B - \theta_i\right) = 0$, where the expectation is with respect to the assumed model (2.1), that is, design-model expectation (Rubin-Bleuer and Schiopu-Kratina, 2005). It follows from (2.2) that more weight is given to the direct estimator $\hat{\theta}_i$ if the model variance σ_v^2 is large relative to the sampling variance ψ_i, and more weight given to the synthetic estimator $\mathbf{z}_i'\boldsymbol{\beta}$ if the sampling variance is large.

The mean squared error (MSE) of the best estimator under the model (2.1) is given by

$$\text{MSE}\left(\tilde{\theta}_i^B\right) = E\left(\tilde{\theta}_i^B - \theta_i\right)^2 = \gamma_i \psi_i, \tag{2.3}$$

where the term $\gamma_i \psi_i$ is often denoted by $g_{1i}(\sigma_v^2)$. It follows from (2.3) that the optimal estimator leads to significant reduction in MSE over the direct estimator if γ_i is small or the model variance is relatively small compared to the total variance $\sigma_v^2 + \psi_i$. This result provides a convincing justification for using the model-based approach to produce small area estimates.

In practice, the model parameters are not known and we replace the parameters in (2.2) by restricted maximum likelihood (REML) estimators $\left(\hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2\right)$ to get the empirical best (EB) estimator:

$$\hat{\theta}_i^{\text{EB}} = \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i)\mathbf{z}_i'\hat{\boldsymbol{\beta}}. \tag{2.4}$$

Rao and Molina (2015), Chapter 6, give details of REML estimation of the model parameters.

2.2 Basic unit level model

We now turn to a basic unit level model which uses unit level sample data $\{(y_{ij}, \mathbf{x}_{ij}),\ j = 1, \ldots, n_i;\ i = 1, \ldots, m\}$, where n_i is the sample size in area i. We assume that the area population means $\bar{\mathbf{X}}_i$ are known. We further assume a basic unit level nested error linear regression model for the population and the same model holds for the sample (Battese, Harter and Fuller, 1988). The sample model is given by

$$y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + v_i + e_{ij}, \quad j = 1, \ldots, n_i; \quad i = 1, \ldots, m, \tag{2.5}$$

where the area random effects $v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$ are assumed to be independent of the unit errors $e_{ij} \overset{\text{iid}}{\sim} N(0, \sigma_e^2)$. Unit level models can lead to significant gains in efficiency over area level models because the model parameters can be estimated more accurately using all the observations in the overall sample, unlike area level models.

For known parameters $(\boldsymbol{\beta}, \sigma_v^2, \sigma_e^2)$, the "best" estimator of the area mean $\bar{Y}_i$ is given by

$$\hat{\bar{Y}}_i^B = E\left[\bar{Y}_i \,\middle|\, \left(y_{ij},\ j = 1, \ldots, n_i;\ \mathbf{x}_{ij},\ j = 1, \ldots, N_i\right), \boldsymbol{\beta}, \sigma_v^2, \sigma_e^2\right] = \bar{\mathbf{X}}_i'\boldsymbol{\beta} + a_i\left(\bar{y}_i - \bar{\mathbf{x}}_i'\boldsymbol{\beta}\right), \qquad (2.6)$$

where $\bar{y}_i$ and $\bar{\mathbf{x}}_i$ are the sample means, $a_i = (1 - f_i)\gamma_i + f_i$ with sampling fraction $f_i = n_i / N_i$ and $\gamma_i = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2 / n_i)$, and N_i is the number of population units in area i (Rao and Molina, 2015, Chapter 7). If the area population size N_i is large and $f_i \approx 0$, then (2.6) reduces to a weighted combination of the "sample regression" estimator $\bar{y}_i + \left(\bar{\mathbf{X}}_i - \bar{\mathbf{x}}_i\right)'\boldsymbol{\beta}$ and the regression synthetic estimator $\bar{\mathbf{X}}_i'\boldsymbol{\beta}$ with weights γ_i and $1 - \gamma_i$ respectively. We denote this approximation to $\hat{\bar{Y}}_i^B$ by $\hat{\mu}_i^B$. As the area sample size n_i increases, the optimal estimator gives more weight to the sample regression estimator. In practice, we replace the model parameters by REML estimators $\left(\hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2, \hat{\sigma}_e^2\right)$ to get the EB estimator $\hat{\bar{Y}}_i^{\text{EB}}$ or $\hat{\mu}_i^{\text{EB}}$.

The EB estimator under the unit level model (2.5) does not account for the survey weights w_{ij}, unlike the area level model. As a result, the EB estimator is not design consistent as the area sample size increases, unless the weights are all equal within the area.

The MSE of $\hat{\mu}_i^B$ is equal to $g_{1i}(\sigma_v^2, \sigma_e^2) = \gamma_i(\sigma_e^2 / n_i)$ while the MSE of the sample regression estimator is equal to σ_e^2 / n_i. It now follows that the optimal estimator leads to significant reduction in MSE over the sample regression estimator if γ_i is small or the model variance σ_v^2 is small relative to the total variance $\sigma_v^2 + \sigma_e^2 / n_i$.

3 Model-based MSE estimators

In this section, we focus on the model-based MSE of EB estimators under the basic area level and unit level models. No closed form expressions for MSE exist, except for a few special cases. This problem has attracted much attention in the SAE literature, leading to second-order approximations to MSE which in turn are used to obtain second-order unbiased estimators of MSE under the assumed models.

3.1 Basic area-level model

We focus on REML estimators of model parameters, denoted $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_v^2$. A second-order unbiased estimator of unconditional model MSE of the EB estimator is given by

$$\text{mse}\left(\hat{\theta}_i^{\text{EB}}\right) = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2g_{3i}(\hat{\sigma}_v^2). \qquad (3.1)$$

Here the leading term in (3.1) is given by (2.3) with σ_v^2 replaced by $\hat{\sigma}_v^2$ and the remaining two terms in (3.1) are of lower order and account for the estimation of $\boldsymbol{\beta}$ and σ_v^2, respectively (see Rao and Molina,

2015, Chapter 6 for details). The MSE estimator (3.1) is positive and second-order unbiased in the sense that its bias is of lower order than $1/m$ for large m. Parametric bootstrap methods have also been used to obtain a MSE estimator. However, the resulting MSE estimator is not second-order unbiased and an additional bias adjustment is made to ensure second-order unbiasedness. Those adjustments typically require double bootstrap methods and some of the adjusted bootstrap MSE estimators may take negative values; see Rao and Molina (2015), Chapter 6.

3.2 Basic unit-level model

We again focus on REML estimation of model parameters in the unit level model (2.5). A positive second-order unbiased estimator of the unconditional MSE of the EB estimator $\hat{\mu}_i^{\text{EB}}$ is given by

$$\text{mse}\left(\hat{\mu}_i^{\text{EB}}\right) = g_{1i}\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right) + g_{2i}\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right) + 2g_{3i}\left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right), \tag{3.2}$$

where the first term is the leading term given in Section 2.2, the second term is due to estimating $\boldsymbol{\beta}$ and the last term is due to estimating σ_v^2 and σ_e^2. The EB estimator $\hat{\mu}_i^{\text{EB}}$ and the associated unconditional MSE estimator (3.2) are valid when the sampling fraction f_i is negligible. We refer the reader to (Rao and Molina, 2015, Section 7.2.3) for MSE estimation in the case of non-negligible sampling fractions.

4 Design MSE estimation

In this section we first study design MSE estimation and then propose composite MSE estimation that provides a balance between the design bias and the coefficient of variation.

4.1 Area-level model

We now turn to estimating the design MSE of the EB estimator by treating the small area parameters θ_i as fixed unknown parameters. As noted in the introduction, survey statisticians are often interested in estimating the design MSE of EB estimators in line with the traditional design MSE estimators of direct estimators for large areas with adequate sample sizes. The design MSE is given by $\text{MSE}_d\left(\hat{\theta}_i^{\text{EB}}\right) = E\left[\left(\hat{\theta}_i^{\text{EB}} - \theta_i\right)^2 \middle| \boldsymbol{\theta}\right]$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)'$ is the vector of area means.

Expressing $\hat{\theta}_i^{\text{EB}}$ as $\hat{\theta}_i + h_i\left(\hat{\boldsymbol{\theta}}\right)$ with $h_i\left(\hat{\boldsymbol{\theta}}\right) = -(1-\hat{\gamma}_i)\left(\hat{\theta}_i - \mathbf{z}_i'\hat{\boldsymbol{\beta}}\right)$, an exactly unbiased estimator of the design MSE is given by

$$\text{mse}_d\left(\hat{\theta}_i^{\text{EB}}\right) = \psi_i + 2\psi_i\left[\partial h_i\left(\hat{\boldsymbol{\theta}}\right)\Big/\partial\hat{\theta}_i\right] + h_i^2\left(\hat{\boldsymbol{\theta}}\right). \tag{4.1}$$

Datta, Kubokawa, Molina and Rao (2011) give an explicit expression for the derivative in the second term of (4.1) in the case of REML estimators of model parameters. The estimator (4.1) can take negative values and can be very unstable in terms of relative root mean squared error (RRMSE) as shown by Datta et al. (2011). It follows that (4.1) is not a reliable estimator of the design MSE, although it is design unbiased.

Our simulation results in Section 5 study the conditional properties of the MSE estimators (3.1) and (4.1) in the design-based framework.

Some theoretical insights can be obtained by focusing on the case of known model parameters and considering the best estimator (2.2) of the area mean θ_i. In this case, Rivest and Belmonte (2000) obtained a design-unbiased estimator given by

$$\text{mse}_d\left(\tilde{\theta}_i^B\right) = \gamma_i \psi_i + (1-\gamma_i)^2\left[\left(\hat{\theta}_i - \mathbf{z}_i'\boldsymbol{\beta}\right)^2 - (\psi_i + \sigma_v^2)\right]. \tag{4.2}$$

Note that for a large sampling variance ψ_i we have $\gamma_i \approx 0$ and (4.2) reduces to

$$\text{mse}_d\left(\tilde{\theta}_i^B\right) \approx \left(\hat{\theta}_i - \mathbf{z}_i'\boldsymbol{\beta}\right)^2 - \psi_i. \tag{4.3}$$

It follows from (4.3) that the MSE estimator can take negative values and, in fact, the probability of getting a negative value is close to 0.5 when γ_i is close to zero or sampling variance ψ_i is large. In this special case of known model parameters, we can study the design bias of the model MSE estimator of (2.2), given by $\text{mse}\left(\tilde{\theta}_i^B\right) = \gamma_i \psi_i$, when averaged over the areas. It can be shown that the average design bias converges in model probability to zero as $m \to \infty$ (Rao and Molina, 2015, page 287). This result suggests that the model MSE estimator should perform well in terms of average design bias, provided the assumed model is valid.

The design-unbiased estimator (4.1) is not usable in practice when it takes a negative value for the sample at hand. Therefore, we propose a modification of (4.1) that leads to a positive MSE estimator. We denote the modified MSE estimator by $\text{mod-mse}_d\left(\hat{\theta}_i^{\text{EB}}\right)$. It uses (4.1) when it takes a positive value for the sample at hand and replaces (4.1) by the model MSE estimate (3.1) when (4.1) takes a negative value. It is possible to use some other positive MSE estimate, for example a naïve positive design-based MSE estimator proposed by Pfeffermann and Gilboa (2017). We have not studied this modification in our simulation study.

We now propose composite estimators of the design MSE that attempt to provide a balance between design bias and RRMSE. One composite estimator is obtained by taking a weighted average of the design MSE estimator (4.1) and the unconditional model MSE estimator (3.1) with weights $\hat{\gamma}_i$ and $(1-\hat{\gamma}_i)$ respectively. This composite MSE estimator may be written as

$$\text{mse}_{c1}\left(\hat{\theta}_i^{\text{EB}}\right) = \hat{\gamma}_i\ \text{mse}_d\left(\hat{\theta}_i^{\text{EB}}\right) + (1-\hat{\gamma}_i)\,\text{mse}\left(\hat{\theta}_i^{\text{EB}}\right). \tag{4.4}$$

It follows from (4.4) that less weight is given to the design MSE estimator when the sampling variance is large and this controls the RRMSE of the composite MSE estimator. Also, the composite MSE estimator has always a smaller design bias than the model MSE estimator. When $\hat{\gamma}_i$ (or the area sample size) is very small, another choice of the compositing weights is to replace $\hat{\gamma}_i$ by $\sqrt{\hat{\gamma}_i}$ and $1-\hat{\gamma}_i$ by $1-\sqrt{\hat{\gamma}_i}$ in (4.4). The resulting composite MSE estimator

$$\text{mse}_{c2}\left(\hat{\theta}_i^{\text{EB}}\right) = \sqrt{\hat{\gamma}_i}\ \text{mse}_d\left(\hat{\theta}_i^{\text{EB}}\right) + \left(1-\sqrt{\hat{\gamma}_i}\right)\text{mse}\left(\hat{\theta}_i^{\text{EB}}\right) \tag{4.5}$$

gives more weight to $\text{mse}_d\left(\hat{\theta}_i^{\text{EB}}\right)$ than (4.4) and thus performs better in terms of design bias at the expense of increased MSE. Similar to (4.4), the alternative composite MSE estimator (4.5) has always a smaller design bias than the model MSE estimator. Both (4.4) and (4.5) can also take on negative values but likely not as often due to their construction. To ensure positive composite MSE estimators, we make a modification similar to $\text{mod-mse}_d\left(\hat{\theta}_i^{\text{EB}}\right)$ and replace (4.4) and (4.5) by the model MSE estimate (3.1) when they take negative values for the sample at hand. We denote the modified estimators by $\text{mod-mse}_{c1}\left(\hat{\theta}_i^{\text{EB}}\right)$ and $\text{mod-mse}_{c2}\left(\hat{\theta}_i^{\text{EB}}\right)$ respectively. In Section 5, we look at the performance of the two modified composite MSE estimators relative to the model MSE estimator (3.1) and the modified design MSE estimator in terms of ARB, RRMSE and coverage rate of confidence intervals.

4.2 Unit-level model

We focus on simple random sampling (SRS) without replacement in each area. Even for this special design, no closed form expressions for the design MSE of the EB estimator $\hat{\bar{Y}}_i^{\text{EB}}$ and its estimator are available in the literature, unlike in the case of the area-level model. Therefore, we propose a heuristic method by evaluating the design MSE of the best estimator $\hat{\bar{Y}}_i^{B}$ given by (2.6), under SRS assuming all the model parameters are known and then estimating the design MSE. The resulting design-unbiased MSE estimator of the best estimator depends on the model parameters and we replace the model parameters by their REML estimators. The resulting MSE estimator is not design-unbiased for the design MSE of the EB estimator and it is likely to underestimate the true design MSE because the variability associated with the estimated model parameters is not taken into account. We study its design performance in a simulation study.

Under SRS without replacement within area i, we have

$$\hat{\bar{Y}}_i^{B} - \bar{Y}_i = a_i\left(\bar{u}_i - \bar{U}_i\right) - (1 - a_i)\bar{U}_i, \tag{4.6}$$

where $\bar{u}_i$ is the area sample mean and $\bar{U}_i$ is the area population mean of the values $u_{ij} = y_{ij} - \mathbf{x}'_{ij}\boldsymbol{\beta}$. It follows from (4.6) that the design MSE of the best estimator is given by

$$\text{MSE}_d\left(\hat{\bar{Y}}_i^{B}\right) = E_d\left(\hat{\bar{Y}}_i^{B} - \bar{Y}_i\right)^2 = a_i^2\, V_d\left(\bar{u}_i\right) + (1 - a_i)^2\, \bar{U}_i^2, \tag{4.7}$$

where

$$V_d\left(\bar{u}_i\right) = n_i^{-1}(1 - f_i)\, S_{ui}^2, \quad \text{and} \quad S_{ui}^2 = (N_i - 1)^{-1}\sum_{j=1}^{N_i}(u_{ij} - \bar{U}_i)^2, \tag{4.8}$$

noting that the cross-product term is zero under SRS.

It now follows from (4.7) and (4.8) that a design unbiased MSE estimator of the best estimator is given by

$$\text{mse}_d\left(\hat{\bar{Y}}_i^{B}\right) = a_i^2\, n_i^{-1}(1 - f_i)\, s_{ui}^2 + (1 - a_i)^2\, \hat{\bar{U}}_i^{2D}, \tag{4.9}$$

where $\hat{\bar{U}}_i^{2D} = n_i^{-1}\sum_{j=1}^{n_i} u_{ij}^2 - N_i^{-1}(N_i - 1)s_{ui}^2$ and $s_{ui}^2 = (n_i - 1)^{-1}\sum_{j=1}^{n_i}\left(u_{ij} - \bar{u}_i\right)^2$. By replacing the model parameters in (4.9) by their REML estimators, a design-based MSE estimator of the EB estimator is obtained, denoted by $\text{mse}_d^*\left(\hat{\bar{Y}}_i^{\text{EB}}\right)$. This MSE estimator is likely to underestimate the design MSE of the EB estimator because the best estimator (2.6) does not account for the variability in the estimators of model parameters.

A composite MSE estimator, $\text{mse}_c^*\left(\hat{\bar{Y}}_i^{\text{EB}}\right)$, is now obtained by taking a weighted combination of $\text{mse}_d^*\left(\hat{\bar{Y}}_i^{\text{EB}}\right)$ and the model-based MSE estimator $\text{mse}\left(\hat{\bar{Y}}_i^{\text{EB}}\right)$ with weights $\hat{\gamma}_i$ and $1 - \hat{\gamma}_i$ respectively. It is given by

$$\text{mse}_c^*\left(\hat{\bar{Y}}_i^{\text{EB}}\right) = \hat{\gamma}_i\,\text{mse}_d^*\left(\hat{\bar{Y}}_i^{\text{EB}}\right) + (1 - \hat{\gamma}_i)\,\text{mse}\left(\hat{\bar{Y}}_i^{\text{EB}}\right). \tag{4.10}$$

Molina and Kominiak (2017) proposed parametric and non-parametric bootstrap estimators of the design MSE of $\hat{\bar{Y}}_i^{\text{EB}}$. They also obtained a composite MSE estimator, similar to (4.10), by using the non-parametric bootstrap (NPB) MSE estimator and the parametric bootstrap (PB) MSE estimator as the two components of the composite MSE estimator associated with $\hat{\gamma}_i$ and $(1 - \hat{\gamma}_i)$ respectively. As noted by the authors, a drawback with this composite MSE estimator is "that it requires to run both PB and NPB procedures for each area, which makes it computationally slower." Molina and Kominiak (2017) also proposed a parametric design bootstrap (PDB) composite MSE estimator. The PDB estimator avoids running both PB and NPB procedures for each area. Both bootstrap composite MSE estimators performed well in a design-based simulation study.

5 Simulation study

In this section, we report the results of limited simulation studies on the design performance of the proposed composite MSE estimators. Section 5.1 gives results for the area level model, and the unit level model results are reported in Section 5.2.

5.1 Area-level model

Following the simulation set up used by Datta et al. (2011), we employ model (2.1) with $m = 30$ areas, $\mathbf{z}_i = (1, z_{i1})'$ for $i = 1, \ldots, m$, where the covariate values $z_{i1}, \ldots, z_{im}$ are generated independently from $N(-1, 1)$ and held fixed over the simulation runs. Further, $\boldsymbol{\beta} = (1, 1)'$, $\sigma_v^2 = 1$ and the sampling variance values are (2.0, 0.6, 0.5, 0.4, 0.2), with each different value of ψ_i assigned to six consecutive areas. Noting that $v_i \sim N(0, 1)$, we generate $\{\theta_i;\ i = 1, \ldots, m\}$ from the linking model $\theta_i = \mathbf{z}_i'\boldsymbol{\beta} + v_i$ and hold them fixed over the simulations to reflect the design-based approach conditioning on the area means θ_i. Then, $R = 100{,}000$ simulated samples $\left\{\hat{\theta}_i^{(r)}:\ i = 1, \ldots, m\right\}$, $r = 1, \ldots, R$ are generated from the sampling model $\hat{\theta}_i = \theta_i + e_i$ with the sampling error e_i generated from $N(0, \psi_i)$ for specified sampling variance ψ_i which is assumed fixed and known. We note that our simulation setup is not exactly design-based but it is "close enough" for the purposes of our study.

From the simulated data $\left\{\left(\hat{\theta}_i^{(r)}, \mathbf{z}_i\right): \ i = 1, \ldots, m\right\}$ the EB estimates $\hat{\theta}_i^{\mathrm{EB}(r)}$ are computed and the MSE of $\hat{\theta}_i^{\mathrm{EB}}$ is approximated by

$$\mathrm{MSE}_i^{\mathrm{EB}} = R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}_i^{\mathrm{EB}(r)} - \theta_i\right)^2. \tag{5.1}$$

The MSE estimators for each simulated sample are computed and averaged over the 100,000 simulation runs. We denote the means of the MSE estimators over the simulations as $\mathrm{mse}_i^{\mathrm{EB}}$, $\mathrm{mse}_{di}^{\mathrm{EB}}$, $\text{mod-mse}_{di}^{\mathrm{EB}}$, $\text{mod-mse}_{c1i}^{\mathrm{EB}}$ and $\text{mod-mse}_{c2i}^{\mathrm{EB}}$, corresponding to model, design unbiased, modified design unbiased, modified composite 1 and modified composite 2 MSE estimators, respectively. The relative bias (RB) of $\mathrm{mse}_i^{\mathrm{EB}}$ is given by

$$\mathrm{RB}_i^{\mathrm{EB}} = \left(\mathrm{mse}_i^{\mathrm{EB}} - \mathrm{MSE}_i^{\mathrm{EB}}\right)/\mathrm{MSE}_i^{\mathrm{EB}} \tag{5.2}$$

where $\mathrm{MSE}_i^{\mathrm{EB}}$ is given by (5.1). The absolute relative bias (ARB) is simply defined as $\mathrm{ARB}_i^{\mathrm{EB}} = \left|\,\mathrm{RB}_i^{\mathrm{EB}}\,\right|$. The terms $\mathrm{ARB}_{di}^{\mathrm{EB}}$, $\mathrm{ARB}_{di-\mathrm{mod}}^{\mathrm{EB}}$, $\mathrm{ARB}_{c1i-\mathrm{mod}}^{\mathrm{EB}}$ and $\mathrm{ARB}_{c2i-\mathrm{mod}}^{\mathrm{EB}}$ are defined in a similar manner.

We also compute the relative root mean squared error (RRMSE) of the MSE estimators over the simulations. We denote those values as $\mathrm{RRMSE}_i^{\mathrm{EB}}$, $\mathrm{RRMSE}_{di}^{\mathrm{EB}}$, $\mathrm{RRMSE}_{di-\mathrm{mod}}^{\mathrm{EB}}$, $\mathrm{RRMSE}_{c1i-\mathrm{mod}}^{\mathrm{EB}}$ and $\mathrm{RRMSE}_{c2i-\mathrm{mod}}^{\mathrm{EB}}$ for the model, design unbiased, modified design unbiased, modified composite 1 and modified composite 2 MSE estimators, respectively. Here RRMSE of the model MSE estimator is defined as

$$\mathrm{RRMSE}_i^{\mathrm{EB}} = \left\{R^{-1}\sum_{r=1}^{R}\left(\mathrm{mse}_i^{\mathrm{EB}(r)} - \mathrm{MSE}_i^{\mathrm{EB}}\right)^2\right\}^{1/2} \Big/ \mathrm{MSE}_i^{\mathrm{EB}}. \tag{5.3}$$

The RRMSE of the other MSE estimators are similarly defined.

We first compare the average over all the areas of $\mathrm{mse}_i^{\mathrm{EB}}$ to the average over all areas of $\mathrm{mse}_{di}^{\mathrm{EB}}$. We obtain 0.42 and 0.35 respectively, showing that the average of the model MSE estimator, 0.42, is close enough to the average of the design MSEs of the EB estimators, 0.35, confirming the theoretical result mentioned in Section 4.1. The theoretical result assumes known model parameters, while the simulation deals with the general case of unknown model parameters.

We next examine the probability of getting a negative value for the three MSE estimators: design unbiased, composite 1 and composite 2. Figure 5.1 shows the percentage of negative values over the simulations for each of the thirty areas. It is clear from Figure 5.1 that the probability of getting a negative value for the design unbiased MSE estimator can be as large as 50% for the first six areas (group 1) with much larger sampling variance relative to the remaining areas (group 2). On the other hand, it is negligible for the areas in group 2. The average probability over areas in group 1 is 45.67% compared to 0.03% in group 2. The probability of getting a negative value for the composite 1 MSE estimator is zero across all thirty areas, while the average probability for the composite 2 MSE estimator is 9.15% over areas in group 1 and zero over areas in group 2. The above results suggest that the composite 1 MSE estimator may not need modification even for areas with large sampling variances. Note that in the current simulation study the composite 1 and modified composite 1 MSE estimators are identical because no zero values were found for the composite 1 MSE estimator.

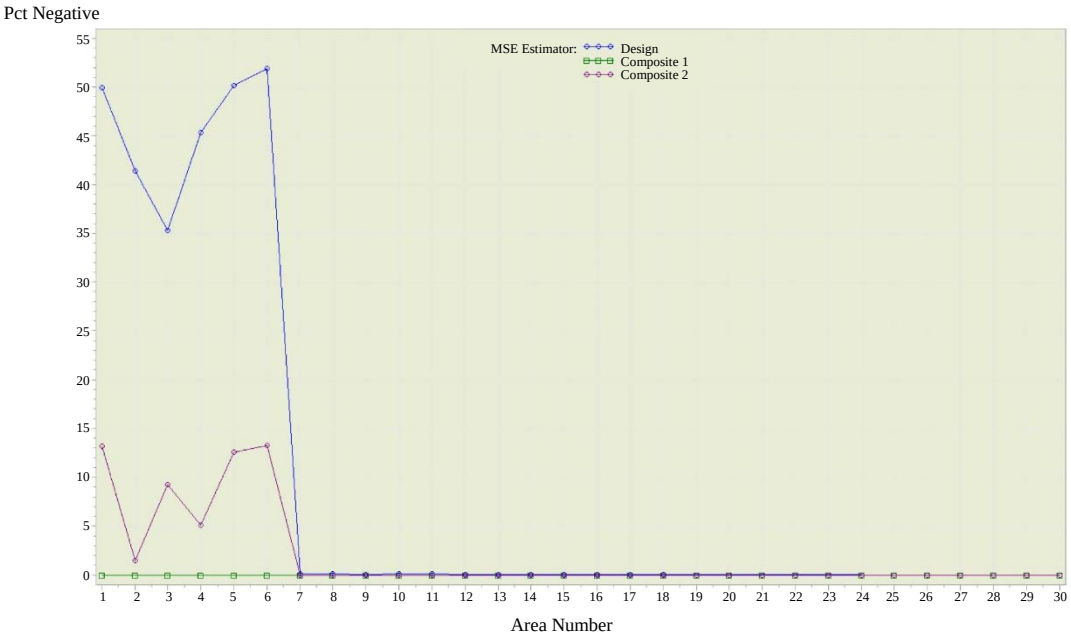

Figure 5.1 Plot of the percent of negative values of the MSE estimators: area level model.

We now turn to the ARB of the MSE estimators. Figure 5.2 shows the ARB values across all the thirty areas for the MSE estimators: model, design unbiased, modified design unbiased, modified composite 1 and modified composite 2 MSE estimators. Table 5.1 gives the mean % design ARB values as well as the mean % design RRMSE over the areas in group 1 and group 2.

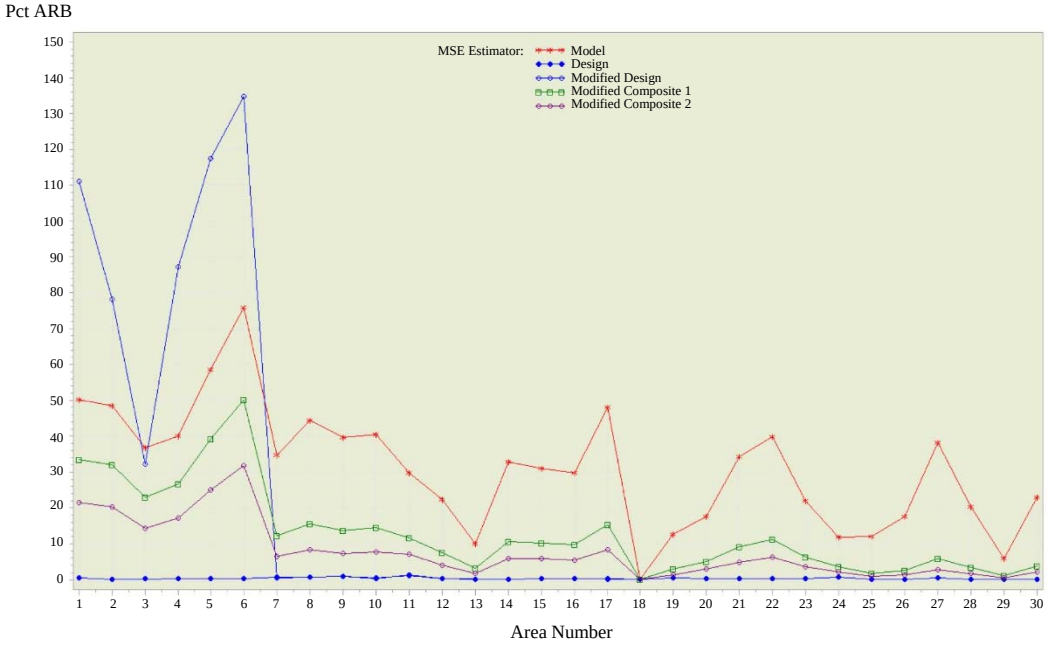

Figure 5.2 Plot of percent ARB of the MSE estimators: area level model.

Table 5.1
Mean % design ARB and mean % design RRMSE of MSE estimators: area level model

MSE Estimator	Mean % design ARB		Mean % design RRMSE	
	Areas 1 to 6	Areas 7 to 30	Areas 1 to 6	Areas 7 to 30
Design	0.33	0.39	246.71	33.62
Modified Design	93.49	0.38	221.86	33.58
Model	51.66	25.76	54.98	26.61
Modified Composite 1	34.08	7.60	96.98	24.70
Modified Composite 2	32.00	4.13	146.31	28.20

As expected, Figure 5.2 shows that the design unbiased estimator has zero ARB (except for simulation errors) across all areas. On the other hand, the modified design unbiased MSE estimator surprisingly exhibits a large ARB for the first six areas, with mean value of 93.49% but negligible for the remaining areas (0.38%). Model MSE estimator also exhibits large ARB for the first six areas with mean ARB of 51.66% that decreases to 25.76% for the remaining areas. On the other hand, the mean ARB for composite 1 MSE estimator is reduced to 34.08% for group 1 and small for group 2 (7.60%). The modified composite 2 MSE estimator that attaches more weight to the design unbiased MSE estimator reduces the mean ARB to 32.00% for group 1 and to 4.13% for group 2.

Figure 5.3 gives a plot of RRMSE of the MSE estimators across all the thirty areas and Table 5.1 reports the mean % RRMSE values for areas in group 1 and group 2. As expected, the design unbiased MSE estimator exhibits very large RRMSE for group 1 with mean value of 246.71%. The modified design unbiased MSE estimator is equally unstable for group 1 (mean RRMSE of 221.86%) in addition to exhibiting large ARB. Model MSE estimator exhibits the smallest RRMSE as expected with mean value of 54.98% for group 1 compared to 96.98% for composite 1 MSE estimator and 146.31% for modified composite 2 MSE estimator. On the other hand, for the areas in group 2 with smaller sampling variances, the mean RRMSE of the three MSE estimators is roughly the same: 24.70% for composite 1, 26.61% for model and 28.20% for modified composite 2 MSE estimators. The mean RRMSE for the design unbiased and modified design unbiased MSE estimators is only slightly larger for group 2 with values of 33.62% and 33.58% respectively.

Finally, we turn to confidence interval coverage rates for a nominal value of 95%. Normal theory coverage rates for the model MSE estimator are computed as

$$\mathrm{CR}\left[\mathrm{mse}\left(\hat{\theta}_i^{\mathrm{EB}}\right)\right] = R^{-1}\sum_{r=1}^{R} I\left[\hat{\theta}_i^{\mathrm{EB}(r)} - 1.96\left(\mathrm{mse}_i^{\mathrm{EB}(r)}\right)^{1/2} \le \theta_i \le \hat{\theta}_i^{\mathrm{EB}(r)} + 1.96\left(\mathrm{mse}_i^{\mathrm{EB}(r)}\right)^{1/2}\right] \quad (5.4)$$

where $I[\bullet]$ is an indicator function with value 1 if θ_i is in the calculated interval and 0 otherwise. Coverage rates for the other MSE estimators are similarly defined. Figure 5.4 is a plot of the percent coverage rates for the MSE estimators. The curve associated with the design-unbiased MSE estimator is not included in the plot because it is not possible to calculate the confidence interval coverage rate due to negative MSE estimates for some simulation runs. Discarding these simulation runs and calculating the intervals from the remaining runs can distort the coverage rate.

The plot shows serious undercoverage for areas in group 1 with large sampling variance. In particular, the mean coverage rate for model, modified composite 1 and modified composite 2 are 68.53%, 78.43%

and 72.87% respectively, whereas the modified design MSE estimator show some improvement: 85.82%. On the other hand, for the areas in group 2 with smaller sampling variances, the mean coverage rate increases to 91.73%, 91.74%, 90.89% and 89.85% for the model, modified composite 1, modified composite 2 and the modified design MSE estimators, respectively. Figure 5.4 suggests that the coverage rates for the model and modified composite MSE estimators are comparable across all areas with the areas in group 1 exhibiting serious undercoverage because of small sample sizes or large sampling variances in those areas.

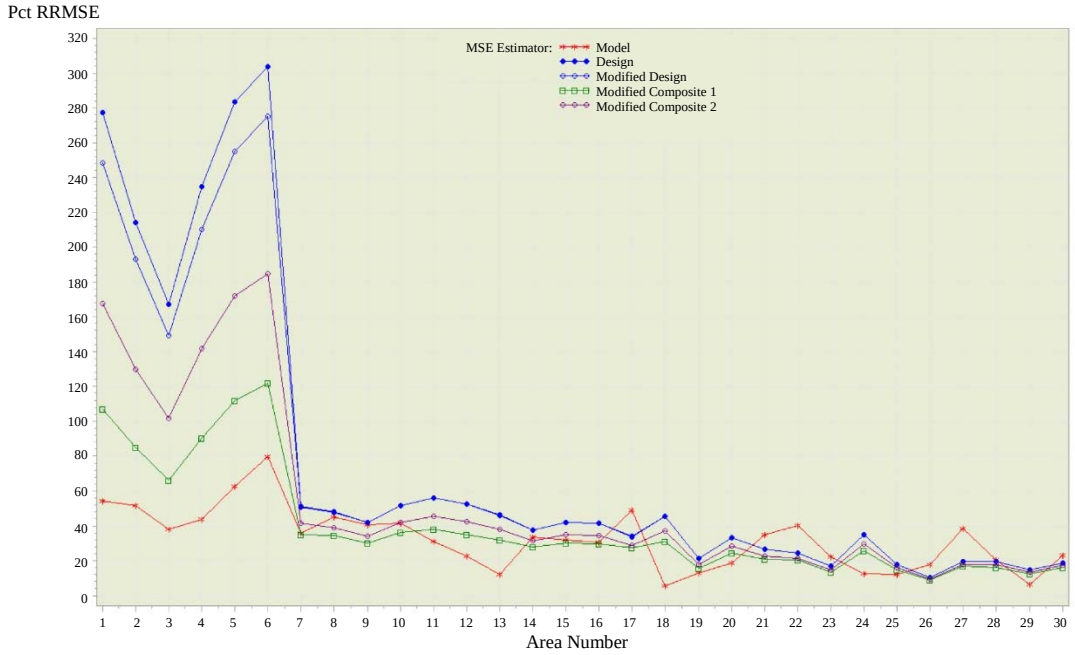

Figure 5.3 Plot of percent RRMSE of the MSE estimators: area level model.

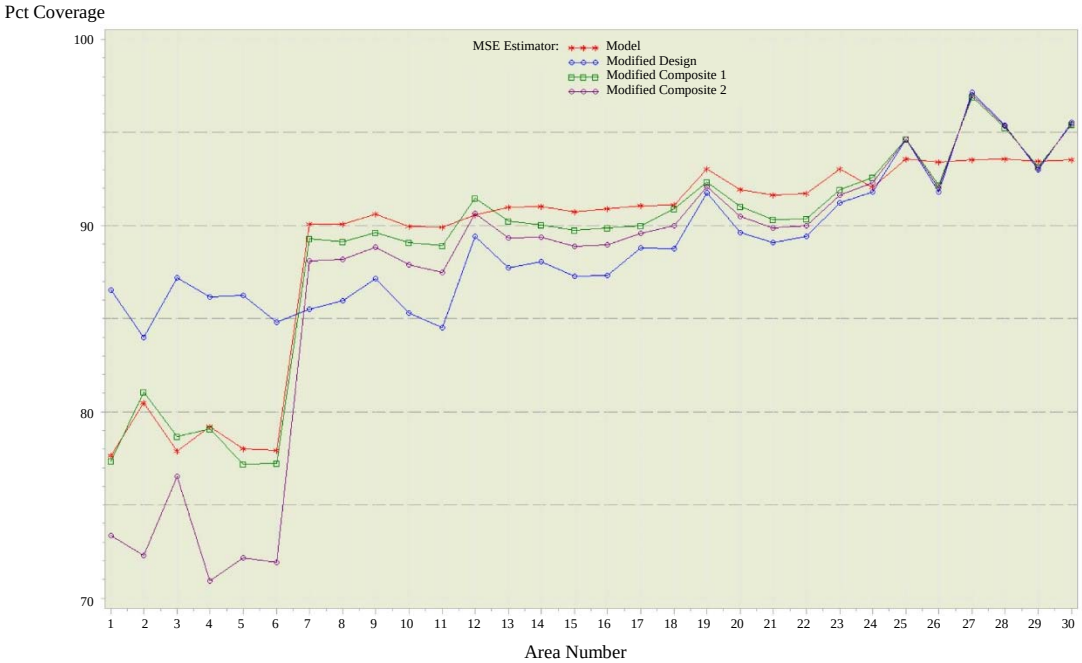

Figure 5.4 Plot of the percent coverage rates for the MSE estimators: area level model.

5.2 Unit-level model

In this section, we report some results of a limited simulation study on the design performance of four MSE estimators under a simple unit-level mean model given by

$$y_{ij} = \beta + v_i + e_{ij}, \quad j = 1, \ldots, N_i; \quad i = 1, \ldots, m \tag{5.5}$$

where the area random effects $v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$ are independent of the unit errors $e_{ij} \overset{\text{iid}}{\sim} N(0, \sigma_e^2)$. The MSE estimators studied include the model MSE estimator $\text{mse}\left(\hat{\bar{Y}}_i^{\text{EB}}\right)$, of the EB estimator $\hat{\bar{Y}}_i^{\text{EB}}$ (Rao and Molina, 2015, Section 7.2.3), the plug-in design-based MSE estimator $\text{mse}_d^*\left(\hat{\bar{Y}}_i^{\text{EB}}\right)$ obtained from (4.9) by replacing the model parameters β, σ_v^2 and σ_e^2 by their REML estimators, the composite MSE estimator given by (4.10), and a "conditional" MSE estimator, $\text{mse}_{\text{CH}}\left(\hat{\bar{Y}}_i^{\text{EB}}\right)$, proposed by Chambers, Chandra and Tzavidis (2001, Section 2.2.2).

For the design-based simulation, we use $m = 30$ small areas and first generate the area population sizes N_i, from a Uniform distribution $U[443, 542]$ and hold them fixed over simulation runs, following Chambers et al. (2011). We generate two fixed finite populations $\{y_{ij},\ j = 1, \ldots, N_i;\ i = 1, \ldots, m\}$ from the mean model (5.5) for specified mean parameter $\beta = 500$ and variance parameters $\sigma_v^2 = 10.40$, $\sigma_e^2 = 94.09$ for the first finite population (denoted Population A) and $\beta = 500$, $\sigma_v^2 = 40.32$, $\sigma_e^2 = 94.09$ for the second finite population (denoted Population B). Note that the variance ratio $\delta = \sigma_v^2 / \sigma_e^2$ is equal to 0.11 for Population A and is smaller than the value 0.43 for Population B. We then draw stratified simple random samples $\{y_{ij},\ j = 1, \ldots, n_i;\ i = 1, \ldots, 30\}$ without replacement, from each finite population, treating each area as a stratum, where the area sample sizes are chosen to be equal: either $n_i = 5$ or $n_i = 20$. In all, we draw $S = 10{,}000$ stratified simple random samples and compute the MSE estimates from each sample. Independently, we also draw $R = 30{,}000$ stratified random samples and compute the EB estimates from each sample. The MSE of the EB estimator for each area is approximated along the lines of (5.1) using the 30,000 simulation runs. Using a large number of simulation runs, $R = 30{,}000$, the true MSE of the EB estimator is accurately approximated by the empirical MSE. On the other hand, a smaller number of simulation runs, such as $S = 10{,}000$, is used for studying the performance of the four MSE estimators to reduce computations. This two-step simulation setup is often used for the unit level model (see e.g., González-Manteiga, Lombardia, Molina, Morales and Santamaria, 2008). Typically, calculating the MSE is much faster than calculating the RB and RRMSE of several MSE estimators, particularly bootstrap MSE estimators.

Using the simulated MSE estimates and the simulated MSE of the EB, we compute the relative bias (RB), the absolute relative bias (ARB) and the relative root mean square error (RRMSE) of the MSE estimators along the lines of (5.2) and (5.3). In the case of Population A and area sample size 5, the plug-in design-based MSE estimator leads to underestimation across all areas, with RB ranging from -87.0% to -18.1%. This underestimation is due to ignoring the variability in the parameter estimates. On the other hand, the model MSE estimator generally overestimates the design MSE with RB ranging from -66.4% to 150.1%. As a result, the composite MSE estimator reduces the underestimation caused by the plug-in

design-based MSE estimator: RB ranging from -55.0% to 115.4%. The conditional MSE estimator overestimates the design MSE consistently with RB ranging from 31.7% to 316.1%. Performance of the MSE estimators in terms of RB improves as the ratio δ increases to 0.43 or the area sample size increases to 20.

Table 5.2 reports the median and mean ARB values for the two populations and the two sample sizes. It shows that the composite MSE estimator performs better than the other MSE estimators for Population A and area sample size 5, with median and mean ARB equal to 53%. On the other hand, the conditional MSE estimator exhibits large median ARB equal to 208% and mean ARB equal to 191%. Median and mean ARB values for all the MSE estimators decrease as the ratio δ increases to 0.43 or the area sample size increases to 20.

Table 5.2
Median and mean % design ARB of MSE estimators: unit level model

MSE Estimator	Population A				Population B			
	$n_i = 5$		$n_i = 20$		$n_i = 5$		$n_i = 20$	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
Design	60.7	54.4	11.2	11.1	8.9	8.9	1.8	2.0
Conditional	207.9	190.7	23.2	19.9	9.4	8.3	0.7	1.0
Model	77.4	81.7	44.0	38.8	29.6	28.4	6.8	8.6
Composite	52.9	53.3	13.1	14.0	7.1	8.8	1.3	1.8

Table 5.3 reports the median and mean % design RRMSE values for the two populations and the two sample sizes. It shows that the model MSE estimator and the composite MSE estimator perform better than the other MSE estimators, especially for Population A and area sample size 5. In the latter case, the plug-in design-based MSE estimator and the conditional MSE estimator exhibit large median and mean RRMSE values: approximately 400% versus 110% for the model MSE estimator and the composite MSE estimator. Performance of all the MSE estimators improves in terms of RRMSE as the ratio δ increases or the area sample size increases. In the case of population B and area sample size 20, model MSE estimator exhibits the smallest median and mean RRMSE: approximately 10% versus 30% for the other MSE estimators.

Table 5.3
Median and mean % design RRMSE of MSE estimators: unit level model

MSE Estimator	Population A				Population B			
	$n_i = 5$		$n_i = 20$		$n_i = 5$		$n_i = 20$	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
Design	414.5	382.0	62.1	60.3	57.6	57.6	29.3	29.0
Conditional	416.5	384.5	64.1	62.2	63.9	64.3	28.4	28.1
Model	107.8	108.5	45.4	41.6	31.6	31.7	8.9	10.8
Composite	113.7	112.9	37.8	38.1	40.7	41.5	26.6	26.4

6 Conclusions

In this paper we studied the properties of alternative MSE estimators in tracking the design MSE of EB estimators of small area means. We examined both area level and unit level models.

In the area level model, we proposed two composite MSE estimators by taking a weighted average of a design unbiased MSE estimator and a model based MSE estimator. Modifications to ensure positive MSE estimators were also given. Performance of the alternative MSE estimators was studied through simulations in terms of absolute relative bias, relative root mean square error and coverage rate of confidence intervals. Our results for the area level model suggest that the design unbiased MSE estimator is not usable in practice when the area sample size is very small because of a large probability of getting a negative value. On the other hand, this probability for the composite 1 MSE estimator (with the same weights as the EB estimator), is either zero or essentially negligible. Our simulations for the area level model for areas with very small sample sizes suggest that the composite 1 MSE estimator leads to smaller ARB relative to the model MSE estimator at the expense of an increase in RRMSE. For areas with larger sample sizes, the ARB of the model MSE estimator persists unlike the ARB of the composite 1 MSE estimator. In terms of coverage rates, the model MSE estimator and the composite 1 MSE estimator are comparable across all areas, but both can lead to serious undercoverage for areas with very small sample sizes. Overall, the composite 1 MSE estimator provides a good compromise in estimating the design MSE.

In the simulation study of the unit level model, our results suggest that the composite MSE estimator generally offers a good compromise between the ARB and RRMSE. However, the plug-in design MSE estimator used in the composite estimator needs modification to take account of the variability in the estimators of model parameters to avoid or reduce the underestimation of design MSE of the EB estimator.

Acknowledgements

The authors thank the assistant and associate editors and the two referees for their comments and suggestions that led to improvements in our examination of the various MSE estimators.

References

Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association,* 83, 401, 28-36.

Beaumont, J.-F., and Bocci, C. (2016). Small area estimation in the Labour Force Survey. Unpublished manuscript.

Chambers, R., Chandra, H. and Tzavidis, N. (2011). On bias-robust mean squared error estimation for pseudo-linear small area estimators. *Survey Methodology*, 37, 2, 153-170. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/2011002/article/11604-eng.pdf.

Datta, G., Kubokawa, T., Molina, I. and Rao, J.N.K. (2011). Estimation of mean squared error of model-based small area estimators. *Test*, 20, 367-388.

Fay, R.E., and Herriot, R.A. (1979). Estimates of income from small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.

González-Manteiga, W., Lombardia, M.J., Molina, I., Morales, D. and Santamaria, L. (2008). Bootstrap mean squared error of a small-area EBLUP. *Journal of Statistical Computation and Simulation*, 78, 443-462.

Molina, I., and Kominiak, E.S. (2017). Estimation of proportions in small areas: Application to the labour force using the Swiss Census Structural Survey. Unpublished technical report.

Pfeffermann, D., and Gilboa, A. (2017). Estimation of randomization MSE in small area estimation. Paper presented at the 2017 SAE conference, Paris.

Rao, J.N.K., and Molina, I. (2015). *Small Area Estimation, Second Edition.* Hoboken, New Jersey: Wiley.

Rivest, L.-P., and Belmonte, E. (2000). A conditional mean squared error of small area estimators. *Survey Methodology*, 26, 1, 67-78. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/2000001/article/5179-eng.pdf.

Rubin-Bleuer, S., and Schiopu-Kratina, I. (2005). On the two-phase framework for joint model and design-based inference. *The Annals of Statistics,* 33, 6, 2789-2810.