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

Survey Methodology

Modified observed best prediction strategies for small area estimation with unit-level data

by Jiangshan Zhang and Jiming Jiang

Release date: December 23, 2025

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 infostats@statcan.gc.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

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, the Agency has developed standards of service which its employees observe in serving its clients. 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.

Published by authority of the Minister responsible for Statistics Canada

© His Majesty the King in Right of Canada, as represented by the Minister of Industry, 2025

Use of this publication is governed by the Statistics Canada Open Licence Agreement.

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, December 2025
Vol. 51, No. 2, pp. 381-398
Statistics Canada, Catalogue No. 12-001-X

Modified observed best prediction strategies for small area estimation with unit-level data

Jiangshan Zhang and Jiming Jiang[1]

Abstract

The observed best prediction (OBP) under a nested-error regression (NER) model was previously proposed using a design-based mean squared prediction error (MSPE) as a tool to derive the best predictive estimator (BPE). A recent study showed the OBP under the NER model may suffer from numerical instability when computing the BPE. We propose several modifications of the OBP under the NER model, including ones using a model-based MSPE to derive the BPE, to improve the numerical stability and predictive performance. We compare the performance of the modified OBP strategies with the existing methods in a simulation study. A real-data example is discussed.

Key Words: Model misspecification; Robustness; Small-area estimation; Unit level.

1. Introduction

Model-based methods of small area estimation (SAE) are typically developed under certain model assumptions. As quoted from George Box, “All models are wrong”. It is quite likely that model misspecification is often inevitable in practical applications. There are, of course, ways to engage more complex models, such as higher-order terms and non-parametric modeling to make the assumed model more flexible. On the other hand, there are practical reasons that a simpler, easier-to-interpret model may be preferred in practice. For example, a linear model is attractive due to its simplicity and ease to interpret, and it utilizes auxiliary variables in a simple way. Note that the auxiliary data may be collected using taxpayer’s money; therefore, it may be “politically incorrect” not to use such information. Statistically, one could run a model selection procedure, which may end up with a more sophisticated model with some of the auxiliary variables eliminated. But, in the end, no one is going to use such a model, practically, due to its complexity and not using information collected via taxpayer’s money. Thus, in such cases, one may not change a model already adopted to use; all one could do is to find a better way of fitting the assumed model so that the resulting small-area predictors are more robust against potential model misspecification.

Jiang, Nguyen and Rao (2011) engaged in such a situation by proposing an SAE method called observed best prediction (OBP) that is remarkably different from the traditional empirical best linear unbiased prediction (EBLUP) method. The idea is to entertain two models, one is the assumed model, which is used to derive the best predictor (BP) when the assumed model holds, and the other is a broader model that is used to derive the parameter estimation under an observe mean squared prediction error (MSPE) measure, leading the the best predictive estimator (BPE). The OBP is the BP with the unknown parameters involved replaced by their BPE. In Jiang et al. (2011), the authors derived the OBP under the area-level Fay-Herriot model (Fay and Herriot, 1979), in which case a model-based MSPE is used to derive the BPE, which

1. Jiangshan Zhang and Jiming Jiang, Department of Statistics, University of California, Davis, CA 95616. E-mail: jimjiang@ucdavis.edu.

assumes that the small-area means are totally unspecified. This is called area-model (AM) OBP here in this paper. The authors showed, both theoretically and empirically, that the AM OBP is more robust than the traditional EBLUP method (e.g., Rao and Molina, 2015) under the same model.

In Jiang et al. (2011), the authors also considered the unit-level nested-error regression (NER) model (Battese, Harter and Fuller, 1988), in which case a design-based MSPE is used to derive the BPE. This is called design-based unit-level (DB/UL) OBP here in this paper. The authors did not study the performance of the proposed DB/UL OBP. Jiang, Nguyen and Rao (2015) showed empirically that the DB/UL OBP is more robust than the EBLUP under the NER model. In a recent study, however, Chen, Lahiri and Salvati (2025) presented some surprising simulation results, which showed the contrary to the finding of Jiang et al. (2015), that is, the DB/UL OBP may perform poorly in terms of the MSPE compared to the EBLUP and some modified OBP methods proposed by Chen et al. (2025). It appears that the numerical stability of the DB/UL OBP is responsible for its poor performance, especially in obtaining the BPE of the variance components under the NER model.

The goal of the current work is two-fold. First, we propose a modification of the DB/UL OBP by imposing limits on the search over the parameter space in the design-based MSPE optimization for computing the BPE. Second, we proposed a model-based unit-level OBP, called MB/UL OBP, under the NER model. Recalled the DB/UL OBP is derived using a design-based MSPE. The new MB/UL OBP is different in that it is based on a model-based MSPE with respect to a broader model that is similar to that of the AM OBP (Jiang et al., 2011; see above). We also propose an alternative approach using plugging-in estimators of the variance components derived from an induced Fay-Herriot model so that the model-based MSPE optimization is only for BPE of the fixed effects.

The methodology developments are described in the next section. In Section 3, we carry out a simulation study that compares the performance of different types of OBP and EBLUP under the NER model, including our proposed modifications and some modifications proposed by Chen et al. (2025). A real-data example is discussed in Section 4.

2. Methods

We first briefly describe the NER model and DB/UL OBP, before proposing several modified OBP strategies under the same model setting.

2.1 The NER model and DB/UL OBP

Consider sampling from finite sub-populations, which are considered samples from some super-populations. Suppose that the sub-populations of outcomes, $\{Y_{ik}, k=1,\dots,N_i\}$, and auxiliary data, $\{X_{ikl}, k = 1,\dots,N_i\}$, $l=1,\dots,p$, are realizations from the corresponding super-populations satisfying the following NER model,

$$Y_{ik} = X'_{ik}\beta + v_i + e_{ik},\ i=1,\dots,m,\ k=1,\dots,N_i, \tag{2.1}$$

where β is a vector of unknown regression coefficients, v_i is a sub-population-specific random effect, and e_{ik} is an error. It is assumed that the v_i's and e_{ik}'s are independent with $v_i \sim N(0, \sigma_v^2)$ and $e_{ik} \sim N(0, \sigma_e^2)$.

Under the finite-population setting, the true small area mean is

$$\theta_i = \bar{Y}_i = \frac{1}{N_i}\sum_{k=1}^{N_i} Y_{ik}$$

for $1 \le i \le m$. Furthermore, write $r_i = n_i / N_i$. Suppose that simple random samples (SRS, e.g., Lohr, 2021) $(y_{ij}, x_{ij}), 1 \le j \le n_i$ are drawn from the finite population, $\{(Y_{ik}, X_{ik}), 1 \le k \le N_i\}, 1 \le i \le m$. Then, a finite-population version of the best predictor (BP), in the sense of minimum MSPE, has the following expression:

$$\tilde{\theta}_i = \mathrm{E}_M(\theta_i \mid y_i) = \bar{X}_i'\beta + \left\{r_i + (1 - r_i)\frac{n_i\rho}{1 + n_i\rho}\right\}(\bar{y}_{i\cdot} - \bar{x}_{i\cdot}'\beta), \tag{2.2}$$

where $y_i = (y_{ij})_{1 \le j \le n_i}$, $\bar{y}_{i\cdot} = n_i^{-1}\sum_{j=1}^{n_i} y_{ij}$, $\bar{x}_{i\cdot} = n_i^{-1}\sum_{j=1}^{n_i} x_{ij}$. Here, E_M denotes the (conditional) expectation under the assumed NER model, β and $\rho = \sigma_v^2 / \sigma_e^2$ are the true parameters under the NER model. Note that the BP is model-dependent.

The next step is to derive the BPE of $\psi = (\beta', \rho)'$. Let $\theta = (\theta_i)_{1 \le i \le m}$ denote the vector of small area means, and $\tilde{\theta} = [\tilde{\theta}_i]_{1 \le i \le m}$ the vector of the BPs, considered as functions of ψ, that is, $\tilde{\theta}_i = \tilde{\theta}_i(\psi)$. The design-based (e.g., Fuller, 2009) MSPE can be expressed as

$$\mathrm{MSPE}(\tilde{\theta}) = \mathrm{E}(|\tilde{\theta} - \theta|^2) = \sum_{i=1}^{m} \mathrm{E}\{\tilde{\theta}_i(\psi) - \theta_i\}^2 = \mathrm{E}\{Q(\psi) + \cdots\}, \tag{2.3}$$

where $\cdots$ does not depend on ψ. Note that the E in (2.3) is different from the E_M in (2.2) in that E is completely model-free; namely, the expectation in (2.3) is with respect to the SRS from the finite populations, which has nothing to do with the assumed model. Furthermore, the function $Q(\psi)$ has the following expression:

$$Q(\psi) = \sum_{i=1}^{m}\left\{\tilde{\theta}_i^2(\psi) - 2\frac{1 - r_i}{1 + n_i\rho}\bar{y}_{i\cdot}\bar{X}_i'\beta + b_i(\rho)\hat{\mu}_i^2\right\} = \sum_{i=1}^{m} Q_i, \tag{2.4}$$

where $b_i(\rho) = 1 - 2a_i(\rho)$ with $a_i(\rho) = r_i + (1 - r_i)\, n_i\rho(1 + n_i\rho)^{-1}$, and Q_i is defined in an obvious way. Furthermore, $\hat{\mu}_i^2$ is a design-unbiased estimator of $\bar{Y}_i^2$ under the SRS, which has the following expression:

$$\hat{\mu}_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{N_i - 1}{N_i(n_i - 1)}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2, \tag{2.5}$$

assuming that $n_i > 1$. The BPE of ψ, $\hat{\psi}$, is the minimizer of $Q(\psi)$ with respect to ψ.

The DB/UL OBP of θ_i is given by (2.2) with ψ replaced by its BPE.

2.2 Restricted OBP for NER model

Although, intuitively, one would expect the DB/UL OBP to be more robust than the EBLUP when the NER model is misspecified, Chen et al. (2025) reported surprising simulation results showing otherwise. It appears that the problem has to do with the numerical optimation of (2.4) over an unrestricted parameter space of ψ.

Note that, although (2.3) holds for any fixed ψ, it may not hold when ψ is random, that is, $\psi = \tilde{\psi}$, which depends on the sampled x and y. However, as it turns out, equation (2.3) still holds approximately if $\tilde{\psi}$ does not vary too much. It was observed empirically that the REML estimator was "always" stable. The numerical instability is due to the BPE under the NER model. This leads to the following modification of the DB/UL OBP.

Let $\tilde{\psi}$ denote the REML estimator of ψ. We propose to optimize (2.4) over a small neighborhood of $\tilde{\psi}$, that is, $\psi_k \in [(1-\delta)\tilde{\psi}_k, (1+\delta)\tilde{\psi}_k]$ for the k the component of ψ, $1 \le k \le p+1$, where $p = \dim(\beta)$, and δ is a small positive number, such as $\delta = 0.05$. This leads to the restricted BPE (RBPE). For the choice of δ, it can be obtained by data-driven techniques. See Section 3.3 for further consideration.

The restricted DB/UL OBP, denoted by RDB/UL OBP, is obtained the same way as the DB/UL OBP except using the RBPE instead of the BPE.

2.3 A model-based OBP for NER model

Jiang et al. (2011) derived the DB/UL OBP under the (SRS) design-based MSPE. Although the latter is almost free of model assumptions, it depends on the sampling design. For example, if stratified sampling (e.g., Lohr, 2021) is used, the DB/UL OBP would not be valid. Of course, one could modify the derivation of the DB/UL OBP to incorporate the sampling design, but this becomes case-by-case, rather than one-formula-for-all.

On the other hand, although a model-based MSPE, such as the one under the broader model in Jiang et al. (2011) to derive the AM OBP (see above), may still involve weak model assumption, it does not depend on the sampling design. In fact, the broader model is almost model-free, whose unit-level extension can be expressed as

$$Y_{ik} = \mu_{ik} + v_i + e_{ik},\ i = 1, \ldots, m,\ k = 1, \ldots, N_i, \tag{2.6}$$

where μ_{ik}'s is the unknown mean of the kth unit in the ith subpopulation, and v_i and e_{ik} are the same as in (2.1). We make the following assumption.

A1. The area-sample mean, $\bar{y}_{i\cdot}$, is a model-unbiased estimator of the area population mean, $\bar{Y}_i = N_i^{-1} \sum_{k=1}^{N_i} Y_{ik}$, that is, $\mathrm{E}(\bar{y}_{i\cdot}) = \mathrm{E}(\bar{Y}_i) \equiv \bar{\mu}_i$, $1 \le i \le m$.

The broader model assumes that each unit has its own unit-level mean, which is very non-restrictive. As for the additional assumption *A1*, it assumes that the area sample mean is an unbiased estimator of the area population mean, which is reasonable. Note that there is a trade-off between model simplicity and robustness. Typically, a more flexible model, such as the broader model assumed above, is more robust to violation of model assumptions (because not much assumption is made, after all). However, the broader model does not utilize information from the auxiliary variables; in this sense, the broader model is not useful in terms of developing predictors of the small area means. The broader model is only used to obtain estimators of the model parameters so that the resulting predictors are more robust to model misspecification. This is the same idea as the OBP (Jiang et al., 2011). On the other hand, a simpler model, such as a linear model, is more attractive to the practitioners; in this sense, it is more useful. Nevertheless a simpler model is more likely to be misspecified, hence less robust to violation of model assumptions.

Under the broader model and *A1*, the small-area mean θ_i can be expressed as

$$\begin{aligned}\theta_i &= \mathrm{E}(\bar{Y}_i \mid v_i) = \frac{1}{N_i}\sum_{k=1}^{N_i}\mathrm{E}(Y_{ik} \mid v_i) = \frac{1}{N_i}\sum_{k=1}^{N_i}\mu_{ik} + v_i \\ &= \frac{1}{N_i}\sum_{k=1}^{N_i}\mathrm{E}(Y_{ik}) + v_i = \mathrm{E}(\bar{Y}_i) + v_i = \mathrm{E}(\bar{y}_{i\cdot}) + v_i.\end{aligned} \tag{2.7}$$

Let $\bar{x}_{i\cdot}^c$ be the mean of $x_{ik}, n_i < k \le N_i$ (recall $1 \le j \le n_i$ correspond to the sample units from the i subpopulation). Then, we have

$$\bar{x}_{i\cdot}^c = \frac{\bar{X}_i}{1-r_i} - \frac{r_i \bar{x}_{i\cdot}}{1-r_i}. \tag{2.8}$$

With (2.8), the BP in (2.2) can be written as

$$\tilde{\theta}_i(\psi) = r_i \bar{y}_{i\cdot} + (1-r_i)\, a_i(\sigma_v^2, \sigma_e^2)\left[\bar{y}_{i\cdot} - \left\{\bar{x}_i - \frac{\bar{x}_{i\cdot}^c}{a_i(\sigma_v^2, \sigma_e^2)}\right\}' \beta\right], \tag{2.9}$$

where $a_i(\sigma_v^2, \sigma_e^2) = n_i \sigma_v^2 / (\sigma_e^2 + n_i \sigma_v^2)$. Under the broader model, we can evaluate MSPE as:

$$\mathrm{MSPE}\{\tilde{\theta}(\psi)\} = \mathrm{E}_{\mathrm{m}}\{\tilde{\theta}(\psi) - \theta\}^2 = \sum_{i=1}^{m}\mathrm{E}_{\mathrm{m}}\{\tilde{\theta}_i(\psi) - \theta_i\}^2 \tag{2.10}$$

where $\tilde{\theta}(\psi) = [\tilde{\theta}_i(\psi)]_{1 \le i \le m}$ and E_{m} denotes the expectation based on the broader model. Substitute (2.9) in (2.10), the MSPE can be expressed as

$$\mathrm{MSPE}\{\tilde{\theta}(\psi)\} = \sum_{i=1}^{m}\mathrm{E}_{\mathrm{m}}\left[r_i \bar{y}_{i\cdot} - \left(\bar{\mu}_i + v_i\right) + (1-r_i)\, a_i\left\{\bar{y}_{i\cdot} - \left(\bar{x}_i - \frac{\bar{x}_{i\cdot}^c}{a_i}\right)' \beta\right\}\right]^2, \tag{2.11}$$

where a_i denotes $a_i(\sigma_v^2, \sigma_e^2)$. Following similar derivations as in Jiang et al. (2011), we have, under assumption *A1* and after some simplification, the following expression:

$$\mathrm{MSPE}\{\tilde{\theta}(\psi)\} = \mathrm{E}_m(I_1 + 2I_2 + I_3), \tag{2.12}$$

where $I_1 = \sum_{i=1}^{m}\left(\bar{y}_i - \bar{\mu}_i - v_i\right)^2$,

$$I_2 = \sum_{i=1}^{m}(1-r_i)\, a_i\left(\bar{y}_i - \bar{\mu}_i - v_i\right)\left[(1-1/a_i)\,\bar{y}_i - (\bar{x}_{i\cdot}' - \bar{x}_{i\cdot}^c / a_i)'\beta\right],$$

and $I_3 = \sum_{i=1}^{m}(1-r_i)^2 a_i^2\{(1-1/a_i)\,\bar{y}_i - (\bar{x}_{i\cdot}' - \bar{x}_{i\cdot}^c/a_i)'\beta\}^2$. As can be seen, I_1 does not depend on β and, under the broader model, we have $\mathrm{E_m}(I_1) = \sum_{i=1}^{m}\sigma_e^2/n_i$. As for I_2, the expectation under the broader model is

$$\begin{aligned}
\mathrm{E_m}(I_2) &= \sum_{i=1}^{m}\mathrm{E}_m\left[(1-r_i)\,a_i\left(\bar{y}_i - \bar{\mu}_i - v_i\right)\left\{(1-1/a_i)\,\bar{y}_i - (\bar{x}_{i\cdot}' - \bar{x}_{i\cdot}^c/a_i)'\beta\right\}\right] \\
&= \sum_{i=1}^{m}(1-r_i)\,a_i\mathrm{E_m}\left[\bar{e}_i\left\{(1-1/a_i)\,\bar{y}_i - (\bar{x}_{i\cdot}' - \bar{x}_{i\cdot}^c/a_i)'\beta\right\}\right] \\
&= \sum_{i=1}^{m}(1-r_i)\,a_i\mathrm{E_m}\left\{\bar{e}_i(1-1/a_i)\,\bar{e}_i\right\} \\
&= \sum_{i=1}^{m}(1-r_i)\,(a_i-1)\,\sigma_e^2/n_i.
\end{aligned} \tag{2.13}$$

It follows that (2.11) can be expressed as

$$\mathrm{MSPE}\{\tilde{\theta}(\psi)\} = \mathrm{E_m}\left(\sum_{i=1}^{m}\left[\frac{\sigma_e^2}{n_i}\{2(1-r_i)\,(a_i-1)+1\} + (1-r_i)^2 a_i^2\left\{\left(1-\frac{1}{a_i}\right)\bar{y}_i - \left(\bar{x}_{i\cdot}' - \frac{\bar{x}_{i\cdot}^c}{a_i}\right)'\beta\right\}^2\right]\right). \tag{2.14}$$

If we reparameterize ψ to be $\psi = (\beta, \rho, \sigma_e^2)$ (recall $\rho = \sigma_v^2/\sigma_e^2$), we can rewrite a_i as $a_i = \rho n_i/(1+\rho n_i)$, which does not depend on σ_e^2. Thus, it can be seen that (2.14) is a monotonically increasing function of σ_e^2. This means that σ_e^2 cannot be estimated by optimizing (2.14). We leave the estimation of σ_e^2 for later. For now, suppose that σ_e^2 is known. Let $\gamma = (\beta, \rho)$. Then, following the OBP idea, we estimate γ by minimizing the expression inside the expectation on the right side of (2.14), that is,

$$\begin{aligned}
Q(\gamma) &= \sum_{i=1}^{m}\left[\frac{\sigma_e^2}{n_i}\{2(1-r_i)(a_i-1)\} + (1-r_i)^2 a_i^2\left\{\left(1-\frac{1}{a_i}\right)\bar{y}_i - \left(\bar{x}_{i\cdot}' - \frac{\bar{x}_{i\cdot}^c}{a_i}\right)'\beta\right\}^2\right] \\
&= \sum_{i=1}^{m}\left[\frac{\sigma_e^2}{n_i}\{2(1-r_i)(a_i-1)\} + \left\{(g_i-1)\,\bar{y}_i - (g_i\bar{x}_{i\cdot} - \bar{X}_i)'\beta\right\}^2\right],
\end{aligned} \tag{2.15}$$

using (2.8), where $g_i = r_i + (1-r_i)\,a_i$. Note that $Q(\gamma)$ is related to the difference between $g_i\bar{x}_{i\cdot}$ and $\bar{X}_i$. Given ρ together with predefined σ_e^2, the β that minimizes $Q(\gamma)$ is

$$\tilde{\beta}\left(\rho, \sigma_e^2\right) = (g_i-1)\,(H_i'H_i)^{-1}H_i\bar{y}_i \tag{2.16}$$

with $H_i = H_i\left(\rho, \sigma_e^2\right) = (g_i\bar{x}_{i\cdot} - \bar{X}_i)$. Denote $H = (H_i)_{1\le i\le m}$ and $L = (g_i\bar{y}_i)_{1\le i\le m}$. We then plug (2.16) into (2.15), and minimize the following over $\rho \ge 0$:

$$\tilde{Q}(\rho) = Q(\gamma)|_{\beta=\tilde{\beta}(\rho,\sigma_e^2)} = \sum_{i=1}^{m}\frac{2\sigma_e^2}{n_i}(1-r_i)\,(a_i-1) + L'\left[I_m - H'(H'H)^{-1}H\right]L. \tag{2.17}$$

Denoted the minimizer by $\tilde{\rho}$. The BPE of β is then obtained by (2.16) with $\rho = \tilde{\rho}$.

As for the case that σ_e^2 is unknown, in practice, it may be estimated using a smoothing technique such as the one demonstrated in Otto and Bell (1995). Here, we recommend using the standard REML estimator of σ_e^2 (e.g., Jiang and Nguyen, 2021, Section 1.3.2) in place of the σ_e^2 in the above derivation.

2.4 Induced area-level model

Chen et al. (2025) also considered an area-level model induced from the NER model:

$$\bar{y}_{i\cdot} = \bar{X}'_{i\cdot}\beta + v_i + \bar{e}_{i\cdot},\ i = 1, \ldots, m, \tag{2.18}$$

where $\bar{e}_{i\cdot} = n_i^{-1} \sum_{j=1}^{n_i} e_{ij}$. The induced model is assumed to satisfy the Fay-Herriot model assumption, where the sampling variance for $\bar{e}_{i\cdot}$ is estimated as $D_i = s^2 / n_i$ and s^2 is the pooled sample variance based on data from all m small areas. The latter authors considered the AM OBP based on (2.18) and showed that it performed better than the DB/UL OBP and the EBLUP in their simulation study.

Here, we propose a different approach by utilizing the unit-level data. First, we use the induced Fay-Herriot model, (2.18), to obtain estimator of σ_v^2. The estimator of σ_e^2 is the same as above, that is, s^2. We then estimate β by minimizing (2.15) over β, where σ_v^2 and σ_e^2 are replaced by these above estimators. We call this method FH/UL OBP.

The motivation for FH/UL OBP is similar to the restricted OBP discussed in Section 2.2. Namely, according to our empirical study, the numerical instability of DB/UL OBP is mainly due to erratic estimates of the variance components; on the other hand, the performance of the β estimator is relatively stable once the variance components are given.

3. Simulation studies

To investigate the performance of the proposed modified OBPs under the NER model as well as their comparison with the existing methods, we consider a simulation setting similar to that of Jiang et al. (2015). To be more specific, we consider a model with a single auxiliary variable, assumed to be linearly associated with the response variable y_{ij} as

$$y_{ij} = \beta_1 x_{ij} + v_i + e_{ij},\ i = 1, \ldots, m,\ j = 1, \ldots, N_i \tag{3.1}$$

where x_{ij} is the value of the auxiliary variable for the *j*th unit of the *i*th area, β_1 is an unknown coefficient associated with the auxiliary variable, v_i and e_{ij} are as in (2.1).

We compared the performance of the following predictors:

(1) The direct estimator (dir);

(2) EBLUP under the NER model (eblup).

(3) the unit-context EBLUP, derived under (2.1) with $X'_{ik}\beta$ replaced by $\beta_1 \bar{X}_i$ (uc-eblup);

(4) the FH/UL OBP (Section 2.4) (fh/ul obp);

(5) AM OBP under the induced Fay-Herriot model (Section 2.4) (fh-obp);

(6) the unit-context OBP, the DB/UL OBP counterpart of (3) (uc-obp);

(7) the MB/UL OBP (Section 2.3) (mb/ul obp);

(8) the RDB/UL OBP (Section 2.2) (rdb/ul obp).

The two unit-context predictors, (3) and (6), in addition to (5), were proposed by Chen et al. (2025).

Below we consider two scenarios of model misspecification to the proposed NER model.

3.1 Complete misspecification

We first consider a complete misspecification situation. The response y for the finite population was generated from the super-population NER model:

$$y_{ij} = b + v_i + e_{ij},\ i = 1,\dots,m,\ j = 1,\dots,N_i. \tag{3.2}$$

The population sizes for all m areas were fixed at $N_1 = \cdots = N_m =$ 1,000. Similarly, the sample sizes for all areas were equally set as $n_i = 4$, and SRS samples were drawn. Two different constant values $b =$ 5, 10 were considered, and simulations were run with the total number of small areas $m =$ 40, 100, or 400. The area-specific effect v_i was generated from $N(0, 1)$, and the error e_{ij} was generated from $N(0, \sigma_i^2)$, where σ_i^2 are independently generated from gamma distribution $\Gamma(3, 0.5)$. The auxiliary variable x for the finite population was generated independently from a log-normal distribution $\text{logNormal}(1, 0.5^2)$. Under each scenario, K = 1,000 replications were performed.

As performance measures, we considered the area-specific and overall empirical MSPEs:

$$\begin{aligned} \text{MSPE}_i &\approx \frac{1}{K}\sum_{k=1}^{K}\left\{\hat{\theta}_i^{(k)} - \theta_i^{(k)}\right\}^2 \\ \text{MSPE} &\approx \frac{1}{K}\sum_{k=1}^{K}\frac{1}{m}\sum_{i=1}^{m}\left\{\hat{\theta}_i^{(k)} - \theta_i^{(k)}\right\}^2, \end{aligned} \tag{3.3}$$

where $\hat{\theta}_i^{(k)}$ and $\theta_i^{(k)}$ are the predicted and true small-area means, respectively, for area i in the kth simulation run. We also consider the overall prediction bias:

$$\text{Bias} = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{m}\sum_{i=1}^{m}\left\{\hat{\theta}_i^{(k)} - \theta_i^{(k)}\right\} \tag{3.4}$$

and the average standard error (SE):

$$\text{SE}_i = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left\{\hat{\theta}_i^{(k)} - \bar{\theta}_i^{(k)}\right\}^2} \quad \text{and} \quad \text{SE} = \frac{1}{m}\sum_{i=1}^{m}\text{SE}_i. \tag{3.5}$$

Table 3.1 reports the simulated overall MSPE, bias, and SE under different numbers of areas and effect size combinations for the case of complete model misspecification, that is, (3.2). There are several observations. In terms of overall MSPE, first, the direct estimators do not change much across simulation scenarios. This is because it is not model-based and the variance component settings were not altered. Second, when $b = 10,$ some predictors can perform worse than the direct estimator; in fact, the EBLUP performed the worst in this case. On the other hand, when $b = 5,$ the direct estimator performed the worst. This suggests that the model helps as long as it is not too severely misspecified. Third, the two unit-context predictors and the two OBPs associated with the Fay-Herriot model performed better than MB/UL OBP and RDB/UL OBP across all simulation scenarios. Among them, the OBP based on the induced Fay-Herriot model has the smallest overall MSPE. Finally, the three unit-level OBPs performed better than the unit-level EBLUP across all scenarios; among the three unit-level OBPs, FH/UL OBP performed the best, followed by MB/UL OBP and the RDB/UL OBP. In fact, in some cases, the FH/UL OBP performed better than the UC-OBP and UC-EBLUP.

Table 3.1
Simulated overall MSPE of predictors: Complete misspecification case

	(m = 40, b = 10)			(m = 100, b = 10)			(m = 400, b = 10)		
Type	**MSPE**	**Bias**	**SE**	**MSPE**	**Bias**	**SE**	**MSPE**	**Bias**	**SE**
dir	1.502	0.005	1.214	1.494	-0.003	1.225	1.494	-0.002	1.223
eblup	1.715	0.353	1.259	1.628	0.245	1.250	1.573	0.186	1.241
uc-eblup	0.823	0.006	0.900	0.793	-0.002	0.890	0.784	-0.001	0.881
fh/ul obp	1.162	0.098	1.081	1.156	0.093	1.080	1.169	0.094	1.078
fh-obp	0.710	0.006	0.840	0.643	-0.002	0.808	0.615	-0.001	0.784
uc-obp	0.813	0.006	0.900	0.792	-0.002	0.890	0.779	-0.001	0.881
mb/ul obp	1.509	0.178	1.247	1.508	0.185	1.239	1.499	0.186	1.220
rdb/ul obp	1.598	0.266	1.222	1.583	0.226	1.239	1.562	0.201	1.224
	(m = 40, b = 5)			**(m = 100, b = 5)**			**(m = 400, b = 5)**		
Type	**MSPE**	**Bias**	**SE**	**MSPE**	**Bias**	**SE**	**MSPE**	**Bias**	**SE**
dir	1.497	0.003	1.229	1.507	0.005	1.219	1.509	0.005	1.226
eblup	1.092	0.475	0.937	1.092	0.496	0.889	1.018	0.498	0.876
uc-eblup	0.812	0.003	0.905	0.784	0.006	0.886	0.773	-0.001	0.880
fh/ul obp	0.831	0.052	0.914	0.770	0.050	0.874	0.749	0.048	0.866
fh-obp	0.777	0.004	0.849	0.648	0.006	0.802	0.613	-0.001	0.784
uc-obp	0.810	0.004	0.906	0.782	0.006	0.886	0.770	-0.001	0.881
mb/ul obp	0.918	0.105	0.956	0.857	0.090	0.915	0.829	0.081	0.881
rdb/ul obp	0.936	0.200	0.952	0.871	0.168	0.914	0.842	0.118	0.907

Note: dir = direct; eblup = empirical best linear unbiased prediction; fh-obp = Fay-Herriot-observed best prediction; fh/ul obp = Fay-Herriot/unit-level observed best prediction; mb/ul obp = model-based/unit-level observed best prediction; MSPE = mean squared prediction error; rdb/ul obp = restricted design-based/unit-level observed best prediction; SE = standard error; uc-eblup = unit-context-empirical best linear unbiased prediction; uc-obp = unit-context-observed best prediction.

In terms of the bias, EBLUP has the highest bias across all scenarios, which is due to the complete misspecification. The Unit-level OBPs have larger bias compared to other unit-context predictors and the Fay-Herriot model based OBPs, but the biases are smaller compared to the EBLUP. FH/UL OBP has the lowest bias among the unit-level OBPs.

Finally, in terms of the SE, the direct estimator, EBLUP, and three unit-level OBPs have higher values compared to the others, with EBLUP having the highest SE when $b = 10.$ However, when $b = 5,$ the direct estimator has the largest SE compared to all the others, and the differences among other predictors are small.

3.2 Partial misspecification

The complete misspecification case may be a little extreme in practice, as the assumed model is completely different from the true underlying model, and such a discrepancy may be detected statistically. Thus, a more realistic situation is considered, where the assumed model is partially correct compared to the true model. In this case, we generate the finite population of y based on the following NER model:

$$y_{ij} = b_0 + b_1 x_{ij} + v_i + e_{ij},\ i = 1,\ldots,m,\ j = 1,\ldots,N_i, \tag{3.6}$$

where $b_0 = b,$ as in the previous case, $b_1 = 4,$ and other settings are the same as in the complete misspecification case. The simulated overall MSPEs are presented in Table 3.2.

Table 3.2
Simulated overall MSPE of predictors: Partial misspecification case

	(m = 40, b = 10)			(m = 100, b = 10)			(m = 400, b = 10)		
Type	**MSPE**	**Bias**	**SE**	**MSPE**	**Bias**	**SE**	**MSPE**	**Bias**	**SE**
dir	9.939	-0.002	3.145	9.914	-0.010	3.141	9.862	-0.003	3.136
eblup	1.752	0.397	1.271	1.637	0.288	1.252	1.582	0.191	1.241
uc-eblup	1.428	-0.001	1.191	1.148	-0.008	1.075	1.148	-0.002	1.011
fh/ul obp	1.568	0.107	1.230	1.170	0.069	1.077	0.969	0.058	0.984
fh-obp	1.598	-0.001	1.256	1.211	-0.009	1.104	1.008	-0.002	1.004
uc-obp	1.427	-0.001	1.190	1.148	-0.009	1.075	1.019	-0.002	1.010
mb/ul obp	1.591	0.314	1.213	1.539	0.267	1.212	1.509	0.195	1.208
rdb/ul obp	1.681	0.334	1.243	1.623	0.267	1.246	1.596	0.199	1.243
	(m = 40, b = 5)			(m = 100, b = 5)			(m = 400, b = 5)		
Type	**MSPE**	**Bias**	**SE**	**MSPE**	**Bias**	**SE**	**MSPE**	**Bias**	**SE**
dir	9.939	0.010	3.158	9.848	0.002	3.141	9.906	-0.004	3.141
eblup	1.109	0.464	0.941	1.041	0.487	0.898	1.016	0.498	0.877
uc-eblup	1.019	0.011	1.191	1.400	0.003	1.068	1.005	-0.004	1.005
fh/ul obp	1.436	0.114	1.195	1.116	0.073	1.049	0.927	0.055	0.963
fh-obp	1.568	0.011	1.261	1.207	0.003	1.090	1.003	-0.004	1.000
uc-obp	1.400	0.011	1.191	1.146	0.003	1.068	1.005	-0.004	1.005
mb/ul obp	1.045	0.465	0.931	0.955	0.477	0.892	0.874	0.475	0.785
rdb/ul obp	1.064	0.341	0.974	0.982	0.320	0.935	0.899	0.241	0.915

Note: dir = direct; eblup = empirical best linear unbiased prediction; fh-obp = Fay-Herriot-observed best prediction; fh/ul obp = Fay-Herriot/unit-level observed best prediction; mb/ul obp = model-based/unit-level observed best prediction; MSPE = mean squared prediction error; rdb/ul obp = restricted design-based/unit-level observed best prediction; SE = standard error; uc-eblup = unit-context-empirical best linear unbiased prediction; uc-obp = unit-context-observed best prediction.

It is seen that, in terms of the overall MSPE, in this case, every model-based predictor clearly outperformed the direct estimator. Furthermore, when $b = 10,$ there seemed to be no clear best performer, although, overall, the two OBPs associated with the Fay-Herriot model (FH/UL OBP and FH-OBP) and the two unit-context predictors performed better than EBLUP, MB/UL OBP and RDB/UL OBP. On the other hand, when $b = 5,$ MB/UL OBP appeared to be the best performer, followed by the RDB/UL OBP.

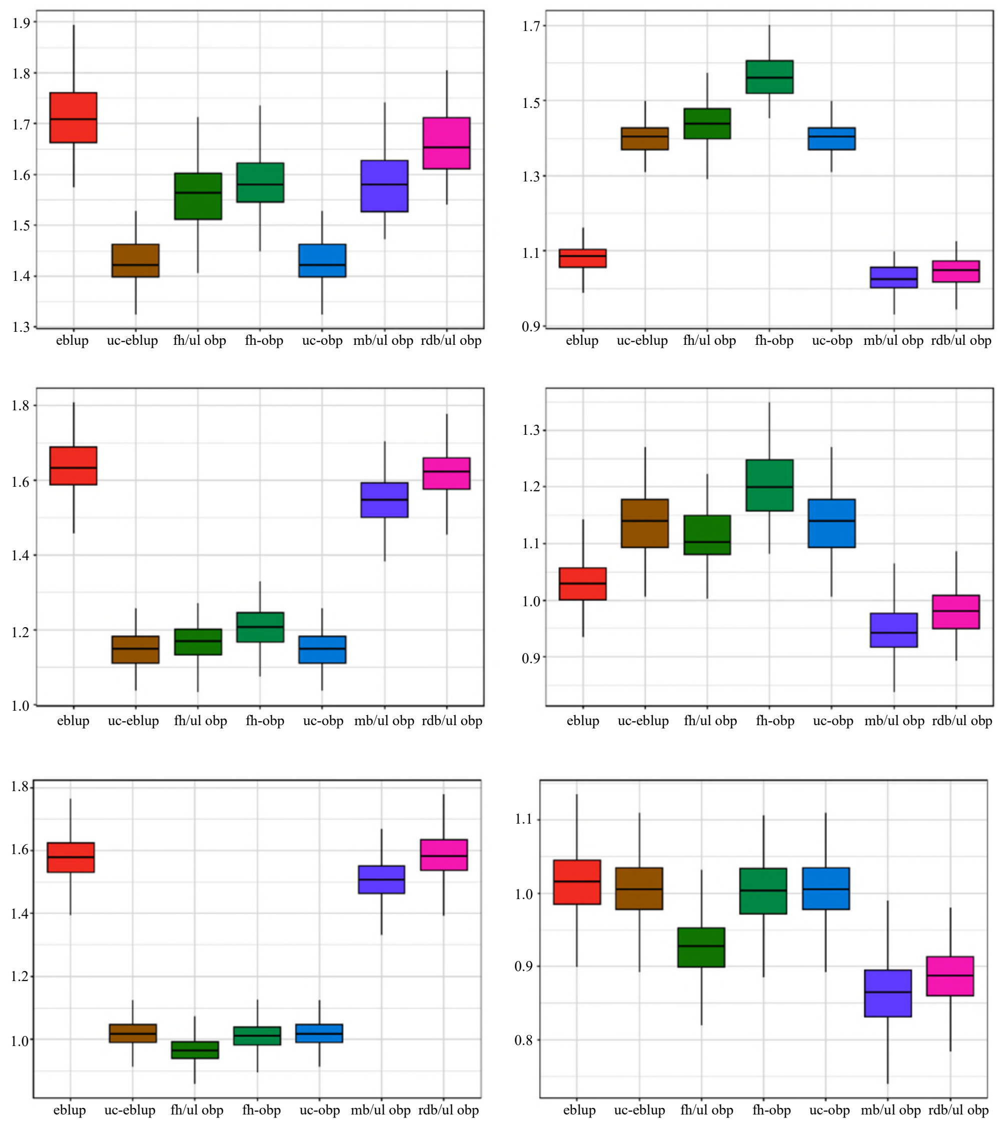
In terms of the bias, still, the unit-context based predictors and direct estimators have smaller bias, while the unit-level based methods now have better performance compared to the complete misspecification case. Once again, EBLUP has the highest bias.

In terms of the SE, the direct estimator has the highest value, as it is not taking any advantage of the model (see below for further discussion). When $b = 10,$ the SEs are not much different across other predictors, while when $b = 5,$ the unit-level based predictors, including the EBLUP, have smaller value compared to unit-context predictors and the two OBPs associated with the Fay-Herriot model.

It is seen that, compared to the complete-misspecification scenario (see Table 3.1), the direct estimator performed much poorer, relative to other estimators, under the partial-misspecification scenario. This is because, under the partial-misspecification scenario, there still exists a linear relationship between Y and X. This allows the model-based methods to take some advantage, while the direct estimator cannot take this advantage.

The boxplots of the area-specific empirical MSPEs of the 7 model-based predictors are shown in Figure 3.1, which provides further detail about their relative performance.

Figure 3.1 Boxplots of empirical area-specific MSPEs: Partial model misspecification



Notes: - Left column: b =10; Right column: b =5. Top: m = 40; Middle: m = 100; Bottom: m = 400.
- eblup = empirical best linear unbiased prediction; fh-obp = Fay-Herriot-observed best prediction; fh/ul obp = Fay-Herriot/unit-level observed best prediction; mb/ul obp = model-based/unit-level observed best prediction; MSPE = mean squared prediction error; rdb/ul obp = restricted design-based/unit-level observed best prediction; uc-eblup = unit-context-empirical best linear unbiased prediction; uc-obp = unit-context-observed best prediction.

Summary: In general, we suggest using the OBP whether model misspecification is present or not. Theoretical and empirical results (e.g., Jiang et al., 2011, 2015) have shown, which is also supported by the trend observed in the current simulation studies, that, in case the underlying model is correctly specified, or almost correctly specified, OBP and EBLUP perform similarly; on the other hand, in case of significant model misspecification, OBP can perform substantially better than EBLUP.

This difference, in fact, suggests a way to diagnose if severe model misspecification exists, which we shall explore in our future work but the basic idea of which is as follows. Let $\hat{\theta}_i$ and $\tilde{\theta}_i$ denote the OBP and EBLUP, respectively, for the ith small area mean, $1 \le i \le m$. It is possible to develop an asymptotic distribution for a standardized version of the squared Euclidean difference, $|\hat{\theta} - \tilde{\theta}|^2 = \sum_{i=1}^{m} (\hat{\theta}_i - \tilde{\theta}_i)^2$, under the null hypothesis that there is no model misspecification. If the null hypothesis is rejected, it would suggest that a severe model misspecification is encountered. In the latter case, OBP would be the preferred method to use; also, in this case, unit-level model based OBPs may not perform as well as the unit-context model or Fay-Herriot model based OBPs.

This difference may also be explored via a data-driven selection of δ with the RDB/UL OBP; see the next subsection. Note that, when $\delta = 0$, OBP is identical to the EBLUP.

Finally, in terms of computational efficiency, we observed no significant differences between the different methods in our simulation studies.

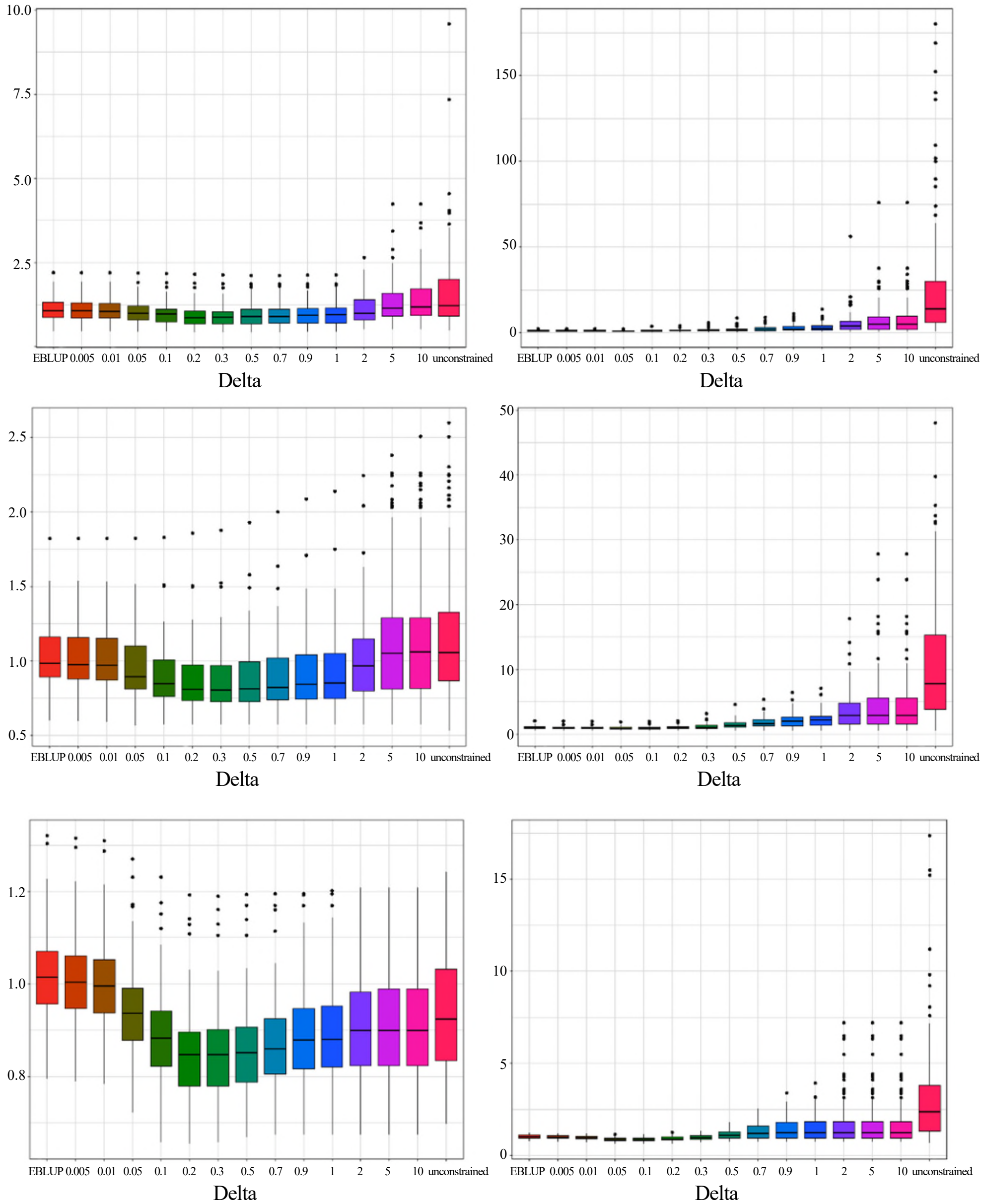
3.3 MSPE of RDB/UL OBP vs δ

To further investigate the relationship between the optimization range, δ, and the performance of RDB/UL OBP, we carried out another simulation study under the same setting of Section 3.2. We consider the constrained DB/UL OBP under different values of δ ranging from 0 to 10, and the unconstrained DB/UL OBP, which corresponds to $\delta = \text{Inf}$. Note that when $\delta = 0$, no optimization is performed, and the result is equal to the initial value, which lead to the EBLUP.

The boxplots of the simulated empirical area-specific MSPEs are presented in Figure 3.2. It shows that, when δ is large, the area-specific MSPEs of RDB/UL OBP is relatively large, and significant outlying values are observed when $\delta = \text{Inf}$. With relatively small, but not too small, values of δ, such as from 0.05 to 0.5, the MSPEs of the RDB/UL OBP are smaller than those of the EBLUP, under all simulation settings. When the model misspecification is moderate, $\delta =$ 0.01 to 0.2 is a good region for resulting smaller MSPE; while for complete misspecification, the optimal range of δ is larger, namely from 0.1 to 0.5. Overall, our simulation results suggest that $\delta =$ 0.1 is a simple choice that ensures improvement of RDB/UL OBP over EBLUP in all cases.

Also, as m increases, the optimal range of δ is closer to 0. The result suggests that, when m is large, a more refined optimization with smaller value of δ is needed.

Figure 3.2 Boxplots of empirical area-specific MSPEs with different range of constraint parameter δ with $b = 5$



Notes: - Left column: Complete model misspecification.; Right column: Partial model misspecification. Top: $m = 40$; Middle: $m = 100$; Bottom: $m = 400$.
- EBLUP = empirical best linear unbiased prediction; MSPE = mean squared prediction error.

4. Real-data example: The TVSFP data

Jiang et al. (2015) used data from the Television School and Family Smoking Prevention and Cessation Project (TVSFP; e.g., Hedeker, Gibbons and Flay, 1994) to illustrate their OBP method under the NER model. Jiang, Rao, Fan and Nguyen (2018) used the same data to demonstrate their proposed classified mixed model prediction method. The original study was designed to test independent and combined effects of a school-based social-resistance curriculum and a television-based program in terms of tobacco use prevention and cessation.

The subjects were seventh-grade students from 28 schools in Los Angeles (LA), California, USA. The students were pretested in January 1986 in an initial study. The same students completed an immediate post-intervention questionnaire in April 1986, a one-year follow-up questionnaire (in April 1987), and a two-year follow-up (in April 1988). Schools were randomized to one of four study conditions: (a) a social-resistance classroom curriculum (CC); (b) a media (television) intervention (TV); (c) a combination of CC and TV conditions; and (d) a no-treatment control. A tobacco and health knowledge scale (THKS) score was one of the primary study outcome variables. The THKS consisted of seven questionnaire items used to assess student tobacco and health knowledge. A student's THKS score was defined as the sum of the items that the student answered correctly. The data is available at www.hsph.harvard.edu/fitzmaur/ala/tvsfp.txt.

The outcome variable for this analysis is the difference between the pretest and immediate post-intervention THKS scores (post-intervention minus pretest). In all, there were 1,600 students from the 28 schools, with the number of students from each school ranging from 18 to 137. Following Jiang et al. (2011), the small areas associated with the 28 LA schools are defined as the 28 subpopulations of students, with each subpopulation matching the characteristics of one of the 28 schools. This way, the data from the 28 schools are viewed as unit-level samples from the 28 small areas. Jiang et al. (2015) considered the following NER model for the TVSFP data:

$$y_{ij} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + v_i + e_{ij}, \tag{4.1}$$

$i = 1, \ldots, 28,\ j = 1, \ldots, n_i,$ where y_{ij} is the unit-level outcome; x_{i1}, x_{i2} are indicators (1 or 0) of CC and TV programs, respectively; $\beta_j, j = 0,1,2,3$ are unknown fixed effects, with β_3 corresponding to the interaction between CC and TV; v_i is an area-specific random effect, and e_{ij} is an additional error. It is assumed that the random effects and errors are independent, with $v_i \sim N(0, \sigma_v^2)$ and $e_{ij} \sim N(0, \sigma_e^2),$ where σ_v^2, σ_e^2 are unknown variances.

Jiang et al. (2015) noted that the proposed NER model is likely misspecified. For example, the outcome is clearly not normal. The authors argued that OBP is appropriate in this situation. Here, we consider the different modifications of the OBP studied in the previous section, and the EBLUP, applied to this data set. The results are presented in Table 4.1 and Table 4.2. The original DB/UL OBP of Jiang et al. (2015), copied from Tables 4.1 and 4.2 of the latter reference, is also included for comparison. It appears that the different methods produced similar predicted values of small area means. In particular, in this case, the FH/UL OBP and FH OBP produced the same results. This is because the auxiliary variables are at the area-level and the ratio n_i / N_i is treated as zero, in which case the two predictors are the same. Due to similar reasons, the MB/UL OBP and RDB/UL OBP are very similar, although not identical. The estimated regression coefficients (fixed effects) by the different methods are presented in Table 4.3. These estimates also seem similar and show the same trend (e.g., positive or negative).

Table 4.1
Predicted small-area means from TVSFP data (Part 1)

ID	CC	TV	eblup	uc-eblup	fh/ul obp	fh-obp	uc-obp	mb/ul obp	rdb/ul obp	db/ul obp
403	1	0	0.913	0.918	0.887	0.887	0.918	0.890	0.890	0.886
404	1	1	0.855	0.835	0.841	0.841	0.844	0.835	0.836	0.844
193	0	0	0.216	0.212	0.215	0.215	0.209	0.213	0.212	0.215
194	0	0	0.220	0.214	0.220	0.220	0.209	0.217	0.217	0.221
196	1	0	0.908	0.915	0.880	0.880	0.918	0.885	0.884	0.878
197	0	0	0.222	0.213	0.223	0.223	0.209	0.219	0.219	0.225
198	1	1	0.809	0.834	0.777	0.777	0.843	0.789	0.787	0.771
199	0	1	0.453	0.449	0.426	0.426	0.452	0.427	0.427	0.426
401	1	1	0.844	0.837	0.825	0.825	0.843	0.824	0.824	0.826
402	0	1	0.200	0.212	0.190	0.190	0.209	0.195	0.194	0.188
405	0	1	0.433	0.450	0.398	0.398	0.452	0.406	0.405	0.394
407	0	0	0.504	0.448	0.500	0.500	0.453	0.480	0.483	0.508
408	1	0	0.904	0.919	0.874	0.874	0.918	0.881	0.880	0.871
409	0	0	0.226	0.215	0.228	0.228	0.209	0.223	0.223	0.230

Note: CC = classroom curriculuml; db/ul obp = design-based/unit-level observed best prediction; eblup = empirical best linear unbiased prediction; fh-obp = Fay-Herriot-observed best prediction; fh/ul obp = Fay-Herriot/unit-level observed best prediction; ID = identification number of schools; mb/ul obp = model-based/unit-level observed best prediction; rdb/ul obp = restricted design-based/unit-level observed best prediction; TV = television; TVSFP = Television School and Family Smoking Prevention and Cessation Project; uc-eblup = unit-context-empirical best linear unbiased prediction; uc-obp = unit-context-observed best prediction.

Table 4.2
Predicted small-area means from TVSFP data (Part 2)

ID	CC	TV	eblup	uc-eblup	fh/ul obp	fh-obp	uc-obp	mb/ul obp	rdb/ul obp	db/ul obp
410	1	1	0.815	0.834	0.783	0.783	0.843	0.794	0.793	0.778
411	0	1	0.444	0.450	0.411	0.411	0.452	0.416	0.416	0.409
412	1	0	0.929	0.916	0.911	0.911	0.919	0.907	0.908	0.913
414	1	0	0.939	0.916	0.926	0.926	0.919	0.918	0.919	0.929
415	1	1	0.870	0.834	0.864	0.864	0.844	0.853	0.854	0.869
505	1	1	0.820	0.834	0.793	0.793	0.843	0.801	0.800	0.790
506	0	1	0.429	0.450	0.393	0.393	0.452	0.402	0.401	0.389
507	0	1	0.452	0.450	0.426	0.426	0.452	0.426	0.427	0.426
508	0	1	0.442	0.449	0.413	0.413	0.452	0.416	0.416	0.411
509	1	0	0.929	0.917	0.913	0.913	0.919	0.910	0.910	0.915
510	1	0	0.906	0.918	0.882	0.882	0.918	0.886	0.885	0.880
513	0	0	0.198	0.213	0.188	0.188	0.209	0.193	0.192	0.185
514	1	0	0.868	0.834	0.862	0.862	0.844	0.851	0.853	0.866
515	0	0	0.193	0.212	0.183	0.183	0.209	0.189	0.188	0.180

Note: CC = classroom curriculuml; db/ul obp = design-based/unit-level observed best prediction; eblup = empirical best linear unbiased prediction; fh-obp = Fay-Herriot-observed best prediction; fh/ul obp = Fay-Herriot/unit-level observed best prediction; ID = identification number of schools; mb/ul obp = model-based/unit-level observed best prediction; rdb/ul obp = restricted design-based/unit-level observed best prediction; TV = television; TVSFP = Television School and Family Smoking Prevention and Cessation Project; uc-eblup = unit-context-empirical best linear unbiased prediction; uc-obp = unit-context-observed best prediction.

Table 4.3
Estimated regression coefficients for TVSFP data

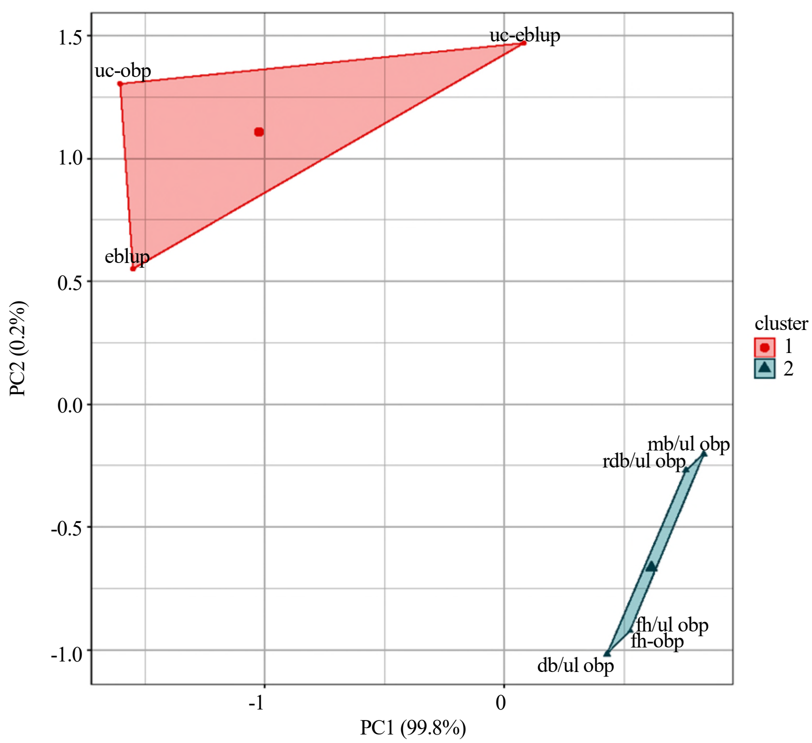
Methods	β_0	β_1	β_2	β_3
eblup	0.211	0.707	0.241	-0.319
uc-eblup	0.212	0.705	0.237	-0.319
fh/ul obp	0.206	0.687	0.214	-0.289
fh-obp	0.206	0.687	0.214	-0.289
uc-obp	0.209	0.709	0.243	-0.318
mb/ul obp	0.206	0.688	0.216	-0.291
rdb/ul obp	0.206	0.689	0.216	-0.292
db/ul obp	0.206	0.687	0.213	-0.288

Note: db/ul obp = design-based/unit-level observed best prediction; eblup = empirical best linear unbiased prediction; fh-obp = Fay-Herriot-observed best prediction; fh/ul obp = Fay-Herriot/unit-level observed best prediction; mb/ul obp = model-based/unit-level observed best prediction; rdb/ul obp = restricted design-based/unit-level observed best prediction; TVSFP = Television School and Family Smoking Prevention and Cessation Project; uc-eblup = unit-context-empirical best linear unbiased prediction; uc-obp = unit-context-observed best prediction.

Despite the apparent similarity, there are some differences when looking deeper into it. Note that we are predicting not one but 28 small-area means, which we need to look at jointly. Furthermore, we are comparing 8 different methods to see if some methods are more similar compared to others when predicting a 28-dimensional vector of small-area means. To facilitate such a comparison, we utilize some modern techniques for high-dimensional data analysis via K-means clustering (K-means) and principal component analysis (PCA). The K-means clustered the 28-dimensional predictors with 2 potential clusters. Here, the number of clusters is determined via the Elbow method (Thorndike, 1953), with the percent of variance explained as the metric. The result classified that EBLUP, UC-EBLUP, and UC-OBP into one class, while FH/UL OBP, FH-OBP, MB/UL OBP, RDB/UL OBP and DB/UL OBP into the other.

To capture the most important differences and similarities among the 28-dimensional predictors, we applied the PCA and then used the top two principal components (PCs), which explain most of the variation in the 28-dimensional predictors, to fit another K-means. The result is shown in Figure 4.1. It is observed that the first PC explained around 99.8% of the variation, while the second PC explained around 0.2% of the variation. This means the top two PCs represent the 28-dimensional predictors very well. The K-means based on the top two PCs resulted in the same clustering result as the above. For the cluster that contains EBLUP, the within-cluster variance is larger compared to the other cluster, although the between-cluster variance explained 84.3% of the total variance.

Figure 4.1 K-means clustering of top 2 principal components of 28-dimensional small area predictors



Note: db/ul obp = design-based/unit-level observed best prediction; eblup = empirical best linear unbiased prediction; fh-obp = Fay-Herriot-observed best prediction; fh/ul obp = Fay-Herriot/unit-level observed best prediction; mb/ul obp = model-based/unit-level observed best prediction; PC = principal components; rdb/ul obp = restricted design-based/unit-level observed best prediction; uc-eblup = unit-context-empirical best linear unbiased prediction; uc-obp = unit-context-observed best prediction.

It may be wondered whether the PCA result is interpretable. We believe the answer depends quite closely on the feature of the specific data structure. Note that, here, we try to compare different methods on their predicted values of a 28-dimensional vector. In general, we try to compare the values of different predictors of an m- dimensional vector of small area means. When m is large, interpretation becomes difficult. In fact, this is a feature of these machine-learning based methods that not everything is interpretable. This not-everything-interpretable feature is nowadays gradually accepted by the data science community, especially if prediction is of primary interest, such as in SAE.

Having said that, as statisticians we always prefer to interpret what is observed. In this occasion, we can see that all but one OBP methods were grouped in one cluster, and all the EBLUP methods, plus the UC-OBP, were grouped in the other. Thus, in some sense, the results suggest that the OBP methods are closer to each other, and the EBLUP methods are closer to each other, which is not surprising. The PCA also grouped the UC-based methods in one cluster, which alse seems reasonable.

In summary, the K-means analysis and PCA revealed additional information about the different small-area predictors under the NER model. While the simulation studies in Section 3 focused on the overall or area-specific MSPEs, the high-dimensional analyses consider the multi-dimensional small-area predictor as a whole, and discover the difference, or similarity, among different methods.

Despite potential differences in small area characteristics, the CC and TV programs appear to effectively improve students' THKS scores. However, whether this translates to better tobacco use prevention and cessation remains in practice unclear. The CC program was relatively more effective than the TV program. Without either intervention, THKS scores showed no improvement in different schools.

Software code. Accessible at https://github.com/Celaeno1017/obpner.

Acknowledgements

The authors are grateful to an Associate Editor and two referees, whose insightful comments have helped improve the work and presentation. Jiming Jiang's research is partially supported by the National Science Foundation of the United States grant DMS-2210569.

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. *J. Amer. Statist. Assoc.*, 80, 28-36.

Chen, Y., Lahiri, P. and Salvati, N. (2025). Effects of model misspecification on small area estimators. *Survey Methodology*, 51(2), 647-657. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2025002/article/00001-eng.pdf.

Fay, R.E. and Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.*, 74, 269-277.

Fuller, W.A. (2009). *Sampling Statistics*, NJ, Hoboken: John Wiley & Sons, Inc.

Hedeker, D., Gibbons, R.D. and Flay, B.R. (1994). Random-effects regression models for clustered data with an example from smoking prevention research. *J. Consulting Clinical Psych.*, 62, 757-765.

Jiang, J. and Nguyen, T. (2021). *Linear and Generalized Linear Mixed Models and Their Applications*. New York: Springer.

Jiang, J., Nguyen, T. and Rao, J.S. (2011). Best predictive small area estimation. *J. Amer. Statist. Assoc.*, 106, 732-745.

Jiang, J., Nguyen, T. and Rao, J.S. (2015). Observed best prediction via nested-error regression with potentially misspecified mean and variance. *Survey Methodology*, 41(1), 37-55. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2015001/article/14200-eng.pdf.

Jiang, J., Rao, J.S., Fan, J. and Nguyen, T. (2018). Classified mixed model prediction. *J. Amer. Statist. Assoc.*, 113, 269-279.

Lohr, S. (2021). *Sampling: Design and Analysis*. Chapman & Hall/CRC.

Otto, M.C. and Bell, W.R. (1995). Sampling error modelling of poverty and income statistics for states. *Proceedings of the Section on Government Statistics*, American Statistical Association, 160-165.

Rao, J.N.K. and Molina, I. (2015). *Small Area Estimation*. New York: John Wiley & Sons, Inc.

Thorndike, R.L. (1953). Who belongs in the family? *Psychometrika*, 18, 267-276.