Statistics and Applications {**ISSN 2452-7395 (online)**}
Volume 16, No. 1, 2018 (New Series), pp 305-314

Observed Best Prediction for Small Area Estimation: A Review

J. Sunil Rao
Division of Biostatistics, University of Miami

Received: March 1, 2018; Revised: May 11, 2018; Accepted: May 20, 2018

Abstract

We provide a distilled review of a relatively new prediction procedure for small area estimation - observed best prediction (OBP) which can be effective under potential model misspecifications. We bring together developments from some of our earlier papers to detail the development of OBP under misspecification of the mean function or more generally, the mean and variance functions. We also briefly review estimation of area-specific MSPEs for these estimators.

Key words. Fay-Herriot model, Mean squared prediction error (MSPE), Model misspecification, Nested-Error regression model, Robustness, Small area estimation (SAE).

1 Introduction

The empirical best linear unbiased prediction, or EBLUP, is well-known in small area estimation (SAE, e.g., Rao 2003, Jiang & Lahiri 2006). There are several ways of deriving the BLUP (e.g., Jiang 2007, pp. 76), but at least one standard procedure is the following. First, one derives the best predictor (BP) of the mixed effects of interests, such as the small area means. Then, one replaces the vector β of the fixed effects by its maximum likelihood estimator (MLE), assuming that the variance components are known (up to this stage one obtains the BLUP). Finally, one replaces the unknown variance components by their ML or REML estimators. It follows that, under the normality assumption, the EBLUP is the BP, in which the unknown fixed parameters, including the fixed effects and variance components, are estimated either by ML or REML. The latter are known to be asymptotically optimal under estimation considerations (e.g., Jiang 2007).

For example, the Fay-Herriot model (Fay and Herriot 1979) is widely used in SAE. The model can be expressed in terms of a mixed effects model:

$$y_i \quad = \quad \mathbf{x}_i'\beta + v_i + e_i, \quad i = 1, \dots, m, \tag{1.1}$$

Corresponding Author: J. Sunil Rao
E-mail: rao.jsunil@gmail.com

where $\mathbf{x}_i$ is a vector of known covariates, β is a vector of unknown regression coefficients, v_i's are area-specific random effects and e_i's are sampling errors. It is assumed that v_i's, e_i's are independent with $v_i \sim N(0, A)$ and $e_i \sim N(0, D_i)$. The variance A is unknown, but the sampling variances D_i's are assumed known. The problem of interest is estimation of the small area means, which, under the assumed model, are the mixed effects $\theta_i = \mathbf{x}_i'\beta + v_i, 1 \leq i \leq m.$

From a practical point of view, any proposed model is subject to model misspecification. Jiang *et al.* (2011) dealt with the case of potential misspecification of the mean function.

Regardless of potential model misspecification, the true small area means should not be dependent on the assumed model. Note that $\theta_i = \mathrm{E}(y_i|v_i), 1 \leq i \leq m$. Now suppose that the true underlying model can be expressed as

$$y_i \quad = \quad \mu_i + v_i + e_i, \quad i = 1, \ldots, m, \tag{1.2}$$

where μ_i's are unknown means, and v_i's and e_i's are the same as in (1). Regardless of the unknown means, $\mathrm{E}(y_i) = \mu_i, 1 \leq i \leq m$. Therefore, under model (2), the small area means can be expressed as

$$\theta_i \quad = \quad \mu_i + v_i \quad = \quad \mathrm{E}(y_i) + v_i, \quad i = 1, \ldots, m. \tag{1.3}$$

The last expression in (3) does not depend on the assumed model. Note that, hereafter, the notation E (without subscript) represents expectation under the true underlying distribution, which is unknown but not model-dependent.

A well-known precision measure for a predictor is the mean squared prediction error (MSPE; e.g., Prasad & Rao 1990, Das, Jiang & Rao 2004). Jiang *et al.* (2011) considered the vector of the small area means $\theta = (\theta_i)_{1 \leq i \leq m}$ and its (vector-valued) predictor $\tilde{\theta} = (\tilde{\theta}_i)_{1 \leq i \leq m}$, the MSPE of the (vector-valued) predictor is defined as

$$\mathrm{MSPE}(\tilde{\theta}) \quad = \quad \mathrm{E}(|\tilde{\theta} - \theta|^2) \quad = \quad \sum_{i=1}^{m} \mathrm{E}(\tilde{\theta}_i - \theta_i)^2. \tag{1.4}$$

Once again, the expectation in (4) is with respect to the true underlying distribution (of whatever random quantities that are involved), which is unknown but <u>not</u> model-dependent. Under the MSPE measure, the BP of θ is its conditional expectation, $\tilde{\theta} = \mathrm{E}(\theta|y)$. Under the assumed model (1), and given the parameters $\psi = (\beta', A)'$, the BP can be expressed as

$$\tilde{\theta}(\psi) \quad = \quad \mathrm{E}_{\mathrm{M},\psi}(\theta|y) \quad = \quad \left[\mathbf{x}_i'\beta + \frac{A}{A + D_i}(y_i - \mathbf{x}_i'\beta)\right]_{1 \leq i \leq m}, \tag{1.5}$$

or, componentwisely, $\tilde{\theta}(\psi)_i = \mathbf{x}_i'\beta + B_i(y_i - \mathbf{x}_i'\beta), 1 \leq i \leq m$, where $B_i = A/(A + D_i)$, and $\mathrm{E}_{\mathrm{M},\psi}$ represents (conditional) expectation under the assumed model with ψ being the true parameter vector. Note that $\mathrm{E}_{\mathrm{M},\psi}$ is different from E unless the assumed model is correct, and ψ is the true parameter vector. Also note that the BP is the minimizer of the area-specific MSPE instead of the overall MSPE (4). In other words, $\tilde{\theta}_i(\psi) = \mathbf{x}_i'\beta + B_i(y_i - \mathbf{x}_i'\beta)$ minimizes $\mathrm{E}(\tilde{\theta}_i - \theta_i)^2$ over all predictor $\tilde{\theta}_i$, if the assumed model (1) is correct and ψ is the true parameter vector. For simplicity, let us assume, for now, that A is known. Then, the precision of $\tilde{\theta}(\psi)$, which is now denoted by $\tilde{\theta}(\beta)$, is measured by

$$\mathrm{MSPE}\{\tilde{\theta}(\beta)\} \;=\; \sum_{i=1}^{m} \mathrm{E}\{B_i y_i - \theta_i + \mathbf{x}_i'\beta(1-B_i)\}^2 \;=\; I_1 + 2I_2 + I_3, \tag{1.6}$$

where $I_1 = \sum_{i=1}^m \mathrm{E}(B_i y_i - \theta_i)^2$, $I_2 = \sum_{i=1}^m \mathbf{x}_i'\beta(1-B_i)\mathrm{E}(B_i y_i - \theta_i)$, $I_3 = \sum_{i=1}^m (\mathbf{x}_i'\beta)^2(1-B_i)^2$. Note that I_1 does not depend on β. As for I_2, by using the expression (3), we have $\mathrm{E}(B_i y_i - \theta_i) = (B_i - 1)\mathrm{E}(y_i)$. Thus, we have $I_2 = -\sum_{i=1}^m (1-B_i)^2 x_i'\beta\mathrm{E}(y_i)$. It follows that the left side of (6) can be expressed as

$$\mathrm{MSPE}\{\tilde{\theta}(\beta)\} \;=\; \mathrm{E}\left\{ I_1 + \sum_{i=1}^m (1-B_i)^2(\mathbf{x}_i'\beta)^2 - 2\sum_{i=1}^m (1-B_i)^2\mathbf{x}_i'\beta y_i \right\}. \tag{1.7}$$

The right side of (7) suggests a natural estimator of β, by minimizing the expression inside the expectation, which is equivalent to minimizing

$$Q(\beta) = \sum_{i=1}^m (1-B_i)^2(\mathbf{x}_i'\beta)^2 - 2\sum_{i=1}^m (1-B_i)^2\mathbf{x}_i'\beta y_i = \beta'\mathbf{X}'\mathbf{\Gamma}^2\mathbf{X}\beta - 2\mathbf{y}'\mathbf{\Gamma}^2\mathbf{X}\beta, \tag{1.8}$$

where $\mathbf{X} = (\mathbf{x}_i')_{1\le i\le m}$, $\mathbf{y} = (y_i)_{1\le i\le m}$ and $\mathbf{\Gamma} = \mathrm{diag}(1-B_i, 1\le i\le m)$. A closed-form solution of the minimizer is obtained as

$$\tilde{\beta} = (\mathbf{X}'\mathbf{\Gamma}^2\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Gamma}^2\mathbf{y} = \left\{\sum_{i=1}^m (1-B_i)^2\mathbf{x}_i\mathbf{x}_i'\right\}^{-1} \sum_{i=1}^m (1-B_i)^2\mathbf{x}_i y_i. \tag{1.9}$$

Here we assume, without loss of generality, that $\mathbf{X}$ is of full column rank. Note that $\tilde{\beta}$ minimizes the "observed" MSPE which is the expression inside the expectation on the right side of (7). Jiang *et al.* (2011) called $\tilde{\beta}$ given by (9) the *best predictive estimator*, or BPE, of β. A predictor of the mixed effects θ is then obtained by replacing β in the BP (5) by its BPE (note that here A is assumed known). This predictor is known as the *observed best predictor*, or OBP.

1. Fay-Herriot model (A unknown). Let us now refer back to the Fay-Herriot model (1) but with A unknown. Again, we begin with the left side of (4), and note that the expectations involved are with respect to the true underlying distribution that is unknown, but not model-dependent. By (5), we have, in matrix expression, $\tilde{\theta}(\psi) = \mathbf{y} - \mathbf{\Gamma}(\mathbf{y} - \mathbf{X}\beta)$, where $\mathbf{\Gamma}$ is defined below (8). By (2) and (3), it can be shown that

$$\mathrm{MSPE}\{\tilde{\theta}(\psi)\} \;=\; \mathrm{E}\{(\mathbf{y}-\mathbf{X}\beta)'\mathbf{\Gamma}^2(\mathbf{y}-\mathbf{X}\beta) + 2A\mathrm{tr}(\mathbf{\Gamma}) - \mathrm{tr}(\mathbf{D})\}, \tag{1.10}$$

where $\mathbf{D} = \mathrm{diag}(D_i, 1\le i\le m)$. The BPE of $\psi = (\beta', A)'$ is obtained by minimizing the expression inside the E on the right side of (10), which is equivalent to minimizing

$$Q(\psi) \;=\; (\mathbf{y}-\mathbf{X}\beta)'\mathbf{\Gamma}^2(\mathbf{y}-\mathbf{X}\beta) + 2A\mathrm{tr}(\mathbf{\Gamma}). \tag{1.11}$$

Let $\tilde{Q}(A)$ be $Q(\psi)$ with $\beta = \tilde{\beta}$ given by (9). It can be shown that $\tilde{Q}(A) = \mathbf{y}'\mathbf{\Gamma}P_{(\mathbf{\Gamma X})^\perp}\mathbf{\Gamma}\mathbf{y} + 2A\mathrm{tr}(\mathbf{\Gamma})$, where for any matrix $\mathbf{M}$, $P_{\mathbf{M}^\perp} = \mathbf{I} - P_{\mathbf{M}}$ with $P_{\mathbf{M}} = \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'$ (assuming nonsingularity of

$\mathbf{M}'\mathbf{M}$), hence, $P_{(\mathbf{\Gamma X})^\perp} = \mathbf{I}_m - \mathbf{\Gamma X}(\mathbf{X}'\mathbf{\Gamma}^2\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Gamma}$ and $\mathbf{I}_m$ is the m-dimensional identity matrix. The BPE of A is the minimizer of $\tilde{Q}(A)$ with respect to $A \geq 0$, denoted by $\tilde{A}$. Once $\tilde{A}$ is obtained, the BPE of β is given by (9) with A replaced by $\tilde{A}$. Given the BPE of ψ, $\tilde{\psi} = (\tilde{\beta}', \tilde{A})'$, the OBP of θ is given by the BP (5) with $\psi = \tilde{\psi}$.

2. Nested-error regression model. Consider sampling from finite subpopulations $P_i = \{Y_{ik}, k = 1, \ldots, N_i\}, i = 1, \ldots, m$. Suppose that auxiliary data $X_{ikl}, k = 1, \ldots, N_i, l = 1, \ldots, p$ are available for each P_i. We assume that the following super-population nested-error regression model (Battese, Harter & Fuller 1988) holds:

$$Y_{ik} = \mathbf{X}_{ik}'\beta + v_i + e_{ik}, \quad i = 1, \ldots, m, \ k = 1, \ldots, N_i, \tag{1.12}$$

where $\mathbf{X}_{ik} = (X_{ikl})_{1 \leq l \leq p}$, the v_i's are small-area specific random effects, and e_{ik}'s are additional errors, such that the random effects and errors are independent with $v_i \sim N(0, \sigma_v^2)$ and $e_{ik} \sim N(0, \sigma_e^2)$. The small area mean for P_i is then $\mu_i = N_i^{-1}\sum_{k=1}^{N_i} Y_{ik}$.

Suppose that $y_{ij}, j = 1, \ldots, n_i$ are observed for the ith subpopulation, $i = 1, \ldots, m$. Let the corresponding auxiliary data be $\mathbf{x}_{ij}, j = 1, \ldots, n_i, i = 1, \ldots, m$. Write $y_i = (y_{ij})_{1 \leq j \leq n_i}$, $y = (y_i)_{1 \leq i \leq m}$, $\bar{y}_{i\cdot} = n_i^{-1}\sum_{j=1}^{n_i} y_{ij}$ and $\bar{\mathbf{x}}_{i\cdot} = n_i^{-1}\sum_{j=1}^{n_i} \mathbf{x}_{ij}$. Let $\psi = (\beta', \sigma_v^2, \sigma_e^2)'$ denote the vector of parameters under the nested-error regression model (12). Under this model with ψ being the true parameter vector, the BP for μ_i is $\mathrm{E}_{\mathrm{M},\psi}(\mu_i|y) = N_i^{-1}\{\sum_{j=1}^{n_i} y_{ij} + \sum_{k \notin I_i} \mathrm{E}_{\mathrm{M},\psi}(Y_{ik}|y_i)\}$, which can be expressed as

$$\tilde{\mu}_i(\psi) = \bar{\mathbf{X}}_i'\beta + \left\{\frac{n_i}{N_i} + \left(1 - \frac{n_i}{N_i}\right)\frac{n_i\sigma_v^2}{\sigma_e^2 + n_i\sigma_v^2}\right\}(\bar{y}_{i\cdot} - \bar{\mathbf{x}}_{i\cdot}'\beta), \tag{1.13}$$

where $\mathrm{E}_{\mathrm{M},\psi}$ denotes the model-based conditional expectation given that ψ is the true parameter vector, I_i is the set of sampled indexes such that Y_{ik} is in the sample iff $k \in I_i$, and $\bar{\mathbf{X}}_i = N_i^{-1}\sum_{k=1}^{N_i} \mathbf{X}_{ik}$ is the subpopulation mean of the $\mathbf{X}_{ik}$'s for the ith subpopulation (which is known).

Note that (13) is a model-based BP, which is not strictly design-unbiased, even under the correct model. Still, the model-based BP is routinely used for SAE under the nested-error regression model due to the anticipated connection between the response and the available covariates (e.g., Rao 2003). Therefore, there is an interest in obtaining estimators of the model parameters that has the best performance in estimating the small area means under the BP operation. The performance of the model-based BP is evaluated by the design-based MSPE. The design-based MSPE is given by

$$\mathrm{MSPE}\{\tilde{\mu}(\psi)\} = \mathrm{E}_{\mathrm{d}}\{|\tilde{\mu}(\psi) - \mu|^2\} = \sum_{i=1}^{m} \mathrm{E}_{\mathrm{d}}\{\tilde{\mu}_i(\psi) - \mu_i\}^2, \tag{1.14}$$

where $\tilde{\mu}(\psi) = [\tilde{\mu}_i(\psi)]_{1 \leq i \leq m}$, $\mu = (\mu_i)_{1 \leq i \leq m}$ and E_{d} denotes the design-based expectation. It can be shown that, assuming simple random sampling within each subpopulation P_i, the MSPE can be expressed as

$$\mathrm{MSPE}\{\tilde{\mu}(\psi)\} = \mathrm{E}_{\mathrm{d}}\left[\sum_{i=1}^{m}\{\tilde{\mu}_i^2(\psi) - 2a_i(\sigma_v^2, \sigma_e^2)\bar{\mathbf{X}}_i'\beta\bar{y}_{i\cdot} + b_i(\sigma_v^2, \sigma_e^2)\hat{\mu}_i^2\}\right], \tag{1.15}$$

where $a_i(\sigma_v^2, \sigma_e^2) = (1 - r_i)\sigma_e^2/(\sigma_e^2 + n_i\sigma_v^2)$, $b_i(\sigma_v^2, \sigma_e^2) = 1 - 2\{r_i + (n_i\sigma_v^2/\sigma_e^2)a_i(\sigma_v^2, \sigma_e^2)\}$ with $r_i = n_i/N_i$, and $\hat{\mu}_i^2$ is a design-unbiased estimator of μ_i^2 given by

$$\hat{\mu}_i^2 \;=\; \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{N_i - 1}{N_i(n_i - 1)}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2 \tag{1.16}$$

Thus, the BPE of ψ is obtained by minimizing

$$Q(\psi) \;=\; \sum_{i=1}^{m}\{\tilde{\mu}_i^2(\psi) - 2a_i(\sigma_v^2, \sigma_e^2)\bar{\mathbf{X}}_i'\beta\bar{y}_{i\cdot} + b_i(\sigma_v^2, \sigma_e^2)\hat{\mu}_i^2\}.$$

A computational procedure for the BPE similar to that for the Fay-Herriot model can be derived as in Jiang *et al.* (2011). Given the BPE $\tilde{\psi} = (\tilde{\beta}', \tilde{\sigma}_v^2, \tilde{\sigma}_e^2)'$, the OBP of μ_i is given by $\tilde{\mu}_i = \tilde{\mu}_i(\tilde{\psi})$, $1 \le i \le m$, where $\tilde{\mu}_i(\psi)$ is given by (13).

2 OBP for Unit Level Model under Misspecification in both the Mean and Variance

Jiang *et al.* (2011) considers misspecification of the mean function, while assuming that the variance-covariance structure of the data is correctly specified. However, the latter, too, may be misspecified in a practical situation. Jiang *et al.* (2015) extend the potential model misspecification to both the mean function and the variance-covariance structure for the unit level model.

Suppose that the subpopulations of responses $\{Y_{ik}, k = 1, \ldots, N_i\}$ and auxiliary data $\{X_{ikl}, k = 1, \ldots, N_i\}, l = 1, \ldots, p$ are realizations from corresponding super-populations that are assumed to satisfy the NER model. It follows that

$$Y_{ik} \;=\; X_{ik}'\beta + v_i + e_{ik},\; i = 1, \ldots, m, k = 1, \ldots, N_i, \tag{2.1}$$

where β, v_i and e_{ik} satisfy the same assumptions as in (12). Under the finite-population setting, the true small area mean is $\theta_i = \bar{Y}_i = N_i^{-1}\sum_{k=1}^{N_i} Y_{ik}$ (as opposed to $\theta_i = \bar{X}_i + v_i$ under the infinite-population setting) for $1 \le i \le m$. Furthermore, write $r_i = n_i/N_i$. Then, the finite-population version of the BP is

$$\tilde{\theta}_i \;=\; \bar{X}_i'\beta + \left\{r_i + (1 - r_i)\frac{n_i\gamma}{1 + n_i\gamma}\right\}(\bar{y}_{i\cdot} - \bar{x}_{i\cdot}'\beta), \tag{2.2}$$

where β and $\gamma = \sigma_v^2/\sigma_e^2$ are the true parameters. The BPE of ψ is the minimizer of

$$Q(\psi) = \sum_{i=1}^{m}\left\{\tilde{\theta}_i^2(\psi) - 2\frac{1 - r_i}{1 + n_i\gamma}\bar{y}_{i\cdot}\bar{X}_i'\beta + b_i(\gamma)\hat{\mu}_i^2\right\} = \sum_{i=1}^{m} Q_i, \tag{2.3}$$

where $\tilde{\theta}_i(\psi)$ is $\tilde{\theta}_i$ with ψ being considered as a parameter vector, $b_i(\gamma) = 1 - 2a_i(\gamma)$ with $a_i(\gamma) = r_i + (1 - r_i)n_i\gamma(1 + n_i\gamma)^{-1}$, and $\hat{\mu}_i^2$ is a design-unbiased estimator of $\bar{Y}_i^2$ given by

$$\hat{\mu}_i^2 \;=\; \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{N_i - 1}{N_i(n_i - 1)}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2 \tag{2.4}$$

(see Jiang *et al.* 2011, sec. 3.2).

3 Estimation of Area-specific MSPE

3.1 Case 1: Misspecification in only the Mean Function

The Prasad-Rao method is well-known in deriving second-order unbiased MSPE estimator for the EBLUP. It is known that the naive estimator of the MSPE (of the EBLUP), which simply replaces the unknown variance components in the (analytic) expression of the MSPE of BLUP by their estimators, underestimates the MSPE of the EBLUP, and is only first-order unbiased if the parameter estimators are consistent. Prasad & Rao (1990) used Taylor expansions to obtain a second-order approximation to the MSPE, and then bias-corrected the plug-in estimator based on the approximation, again to the second-order, to obtain an estimator of the MSPE whose bias is $o(m^{-1})$, where m is the number of small areas. The method has since been used extensively in SAE and several extensions have been given (e.g., Lahiri & Rao 1995, Datta & Lahiri 2000, Jiang & Lahiri 2001, Datta, Rao & Smith 2005). The Prasad-Rao method is based on the assumption that the underlying model is correct, hence the existence of the true parameters. In our case, however, such an assumption is not made, which makes estimation of the area-specific MSPE much more difficult.

Jiang *et al.* (2011) derived a second-order unbiased estimator of the MSPE of $\tilde{\theta}_i$ but unlike the P-R MSPE estimator, their estimator was not guaranteed to be nonnegative. As an alternative, they used the following bootstrap method to obtain an MSPE estimator that is guaranteed nonnegative. The development in Jiang *et al.* (2011) focused on the F-H model. Let $\tilde{\theta} = (\tilde{\theta}_i)_{1\leq i\leq m}$ denote the vector of OBP. First generate $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(L)}$ independently from the $N(\tilde{\theta}, \mathbf{D})$ distribution, where $\mathbf{D} = \mathrm{diag}(D_i, 1 \leq i \leq m)$. Let $\tilde{\theta}_i^{(l)}$ be the OBP for θ_i based on $\mathbf{y}^{(l)}$, $1 \leq i \leq m, 1 \leq l \leq L$. Then, the bootstrap estimator of $\mathrm{MSPE}(\tilde{\theta}_i)$ is given by

$$\widetilde{\mathrm{MSPE}}_{\mathrm{b}}(\tilde{\theta}_i) \quad = \quad \frac{1}{L}\sum_{l=1}^{L}\{\tilde{\theta}_i^{(l)} - \tilde{\theta}_i\}^2. \tag{3.1}$$

3.2 Case 2: Misspecification in both the Mean and Variance Functions

Jiang *et al.* (2015) used a technique known as partial derivation to derived $\widehat{\mathrm{MSPE}}(\hat{\theta}_i)$, a second-order unbiased estimate but showed that it is not guaranteed nonnegative (for technical details, the reader is referred to the above reference). Note that the problem of negative MSPE estimates is also encountered in Jiang *et al.* (2011), where the authors propose to substitute the negative estimate with a bootstrap MSPE estimate, which is guaranteed nonnegative. Jiang *et al.* (2011) used a parametric bootstrap method to obtain the MSPE estimator, although the justification is questionable given the potential model misspecification. In Jiang *et al.* (2015), they proposed to use the nonparametric bootstrap following Efron's original idea (Efron 1979). The method does not rely on the NER model, hence is not affected by the model misspecification.

Suppose that the small area subpopulations, or the N_i's, are large enough, so that the sampling from the subpopulations can be treated approximately as with replacement. Let $z_{ij} = (x'_{ij}, y_{ij})', j = 1, \dots, n_i$ denote the samples from the ith small area, $1 \le i \le m$. We then draw samples, $z_{ij}^{(b)} = [\{x_{ij}^{(b)}\}', y_{ij}^{(b)}]', j = 1, \dots, n_i$, with replacement, from $\{z_{ij}, j = 1, \dots, n_i\}$, independently for $1 \le i \le m$. Suppose that B bootstrap samples are drawn, resulting $z^{(b)} = \{z_{ij}^{(b)}, 1 \le j \le n_i, 1 \le i \le m\}, 1 \le b \le B$. Let $\hat{\theta}_i^{(b)}$ denote the OBP of $\theta_i^{(b)} = \bar{y}_{i\cdot}$ based on $z^{(b)}$. For example, the bootstrap version of (18) is

$$\tilde{\theta}_i^{(b)} = \bar{x}'_{i\cdot}\beta + \left\{ r_i + (1 - r_i)\frac{n_i\gamma}{1 + n_i\gamma} \right\} [\bar{y}_{i\cdot}^{(b)} - \{\bar{x}_{i\cdot}^{(b)}\}'\beta].$$

Then, the bootstrap estimator of $\mathrm{MSPE}(\hat{\theta}_i) = \mathrm{E}(\hat{\theta}_i - \bar{Y}_i)^2$ is

$$\widehat{\mathrm{MSPE}}_{\mathrm{b}}(\hat{\theta}_i) = \frac{1}{B}\sum_{b=1}^{B}\{\hat{\theta}_i^{(b)} - \bar{y}_{i\cdot}\}^2. \tag{3.2}$$

It is clear that $\widehat{\mathrm{MSPE}}_{\mathrm{b}}(\hat{\theta}_i)$ is always nonnegative. On the other hand, the bootstrap MSPE estimator is not second-order unbiased.

4 Television School and Family Smoking Prevention and Cessation Project (TVSFP)

To illustrate the performance of the OBP and it's MSPE estimation, we review an analysis of the TVSFP data as done in Jiang *et al.* (2015). 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 the Los Angeles (LA) and San Diego, California, areas. The students were pretested in January 1986 in an initial study. The same students completed an immediate postintervention questionnaire in April 1986, a one-year follow-up questionnaire (in April 1987), and a two-year follow-up (in April 1988). Jiang *et al.* (2015) considered a subset of the TVSFP data involving students from 28 LA schools, where the 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, and the one used for this analysis. Only data from the pretest and post-intervention are available for the current analysis. More specifically, the data only involved subjects who had completed the THKS at both of these time points. In all, there were 1,600 students from the 28 schools, with the number of students from each school ranging from 18 to 137.

Jiang *et al.* (2015) considered the problem of estimating the small area means for the difference between the post-intervention and pretest THKS scores (the response). Here the "small area" was understood to be the number of major characteristics (e.g., residential area, teach/student ratio) that affect the response, but are not captured by the covariates in the model (i.e., linear combination of the CC, TV and CCTV indicators). Thus, a small area is the seventh graders in all of the U.S. schools that share the similar major characteristics as a LA school involved in the data over a reasonable period of time (e.g., 5 years). There were 28 LA schools in the TVSFP data that correspond to 28 sets of characteristics, so that the data are considered random samples from the 28 small areas defined as above. As such, each small area population is large enough so that $n_i/N_i \approx 0, 1 \leq i \leq 28$. Recall that the n_i's in the TVSFP sample range from 18 to 137, while the N_i's are expected to be at least tens of thousands. Note that the only place in the OBP where the knowledge of N_i is required is through the ratio n_i/N_i. The proposed NER model can be expressed with $x'_{ij}\beta = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,1} x_{i,2}$, where $x_{i,1} = 1$ if CC, and 0 otherwise; $x_{i,2} = 1$ if TV, and 0 otherwise. It follows that all the auxiliary data x_i are at the area level; as a result, the value of $\bar{X}_i$ is known for every i.

It should be noted that the response, y_{ij}, is difference in the THKS scores, whose possible values are integers between 0 and 7. Clearly, such data is not normal. The potential impact of the non-normality is two-fold. On the one hand, it is likely that the original NER model, as proposed by Hedeker *et al.* (1994), is misspecified. On the other hand, even without the normality, the best linear predictor can still be justified (BLP; e.g., Searle, Casella & McCulloch 1992, sec. 7.3). Furthermore, the Gaussian ML (REML) estimators are consistent and asymptotically normal even without the normality assumption (Jiang 1996; also see Jiang 2007, ch. 1). Other aspects of the NER model include homoscedasticity of the error variance across the small areas. Jiang *et al.* (2015) showed that a bimodal shape suggesting potential heteroscedasticity in the error variance.

The OBP analysis for the first 14 of the 28 school small areas are presented in Table 1. Jiang *et al.* (2015) also report results for the remainder of the school small areas. Also presented are the corresponding $\widehat{\mathrm{MSPE}}$ and $\widehat{\mathrm{MSPE}}_{\mathrm{b}}$, and their square roots as the measures of uncertainty. Out of the 14 small areas, the value of $\widehat{\mathrm{MSPE}}$ is negative for 4 of them; hence the square roots are not available. However, the values of $\widehat{\mathrm{MSPE}}_{\mathrm{b}}$ are all positive. It is seen that the OBPs are all positive, even for the small areas in the control group. There were 7, 8, 7 and 7 small areas in the (0,0), (0,1), (1,0) and (1,1) groups, respectively.

In conclusion, Jiang *et al.* (2015) concluded in spite of the potential difference in the small area characteristics, the CC and TV programs appears to be successful in terms of improving the students' THKS scores It also seems apparent that the CC program was relatively more effective than the TV program. Without the intervention of any of these programs, the THKS score did not seem to improve in terms of the small area means.

5 Discussion

The idea of OBP may be viewed more broadly as dealing with two models. The first is a broader model under no or very weak assumptions. Such models are robust against model misspecifications. However, the broader model is less useful in terms of addressing the practical interest. For example, it is often desirable to make use of covariate data that are available from the surveys, and such information would be more useful if the association between the response and covariates is relatively simple. This brings in the second model which is more restrictive, but relatively simple and explores the explicit associations in variables. The OBP approach is to develop a prediction

Table 1: **OBP, Measures of Uncertainty: Schools 1–14**

ID	CC	TV	OBP	$\sqrt{\widehat{\text{MSPE}}_b}$	$\sqrt{\widehat{\text{MSPE}}}$
403	1	0	.886	.171	.137
404	1	1	.844	.296	*
193	0	0	.215	.207	.268
194	0	0	.221	.137	.332
196	1	0	.878	.171	.298
197	0	0	.225	.158	.337
198	1	1	.771	.220	*
199	0	1	.426	.142	*
401	1	1	.826	.133	*
402	0	0	.188	.171	.358
405	0	1	.394	.147	.143
407	0	1	.508	.300	.364
408	1	0	.871	.240	.344
409	0	0	.230	.125	.344

method based on the second model that is more robust against misspecification of the underlying model, and we do this by measuring the performance of the predictor based on the first model. The OBP idea has continued to see further developments. For instance, Chen *et al.* (2015) have extended OBP for small area count data. Jiang *et al.* (2018) have also used OBP in a new mixed model prediction paradigm known as classified mixed model prediction (CMMP) which is intended for situations where the group membership of future areas is not known. More work needs to be done on further improving area-specific MSPE estimates - in particular the bootstrapping strategies that we have summarized here.

Acknowledgements. J. Sunil Rao is partially supported by NSF grant DMS-1513266 and NIH grants R01-CA160593A1 and U54-MD010722.

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 American Statistical Association*, **80**, 28-36.

Chen, S., Jiang, J. and Nguyen, T. (2015). Observed best prediction for small area counts, *Journal of Survey Statistics and Methodology*, **3**, 136-161.

Das, K., Jiang, J. and Rao, J. N. K. (2004), Mean squared error of empirical predictor, *Annals of Statistics*, **32**, 818-840.

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

Datta, G. S., Rao, J. N. K. and Smith, D. D. (2005), On measuring the variability of small area estimators under a basic area level model, *Biometrika*, **92**, 183-196.

Efron, B. (1979), Bootstrap method: Another look at the jackknife, *Annals of Statistics*, **7**, 1–26.

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

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

Jiang, J. (1996), REML estimation: Asymptotic behavior and related topics, *The Annals of Statistics*, **24**, 255–286.

Jiang, J. (2007), *Linear and Generalized Linear Mixed Models and Their Applications*, Springer, New York.

Jiang, J. and Lahiri, P. (2006), Mixed model prediction and small area estimation (with discussion), *TEST*, **15**, 1-96.

Jiang, J., Nguyen, T. and Rao, J. S. (2011), Best predictive small area estimation, *Journal of American Statistical Association*, **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**, 37-55.

JIang, J., Rao, J.S., Fan, J. and Nguyen, T. (2018). Classified mixed model prediction, *Journal of American Statistical Association*, (to appear).

Lahiri, P. and Rao, J. N. K. (1995), Robust estimation of mean squared error of small area estimators, *Journal of American Statistical Association*, **90**, 758-766.

Prasad, N. G. N. & Rao, J. N. K. (1990), The estimation of mean squared errors of small area estimators, *Journal of American Statistical Association*, **85**, 163-171.

Rao, J. N. K. (2003), *Small Area Estimation*, Wiley, New York.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992), *Variance Components*, Wiley, New York.