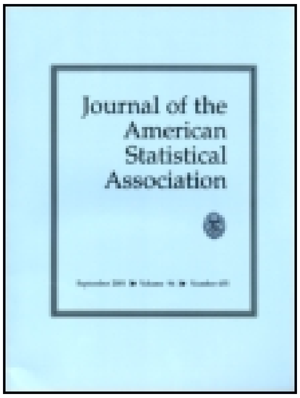

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

Best Predictive Small Area Estimation

Jiming Jiang[a], Thuan Nguyen[a] & J. Sunil Rao[a]

[a] Jiming Jiang is Professor, Department of Statistics, University of California, Davis, Davis, CA 95616. Thuan Nguyen is Assistant Professor, Department of Public Health and Preventive Medicine, Oregon Health and Science University, Portland, OR 97239-3098. J. Sunil Rao is Professor, Department of Epidemiology and Public Health, University of Miami, Miami, FL 33136. Jiming Jiang is partially supported by the NSF grant DMS-0809127. J. Sunil Rao is partially supported by the NSF grant DMS-0806076. The research of all three authors are partially supported by the NIH grant R01-GM085205A1. The authors are grateful to an Associate Editor and two Referees for their constructive comments that have led to improvements of the article.
Published online: 24 Jan 2012.

Best Predictive Small Area Estimation

Jiming JIANG, Thuan NGUYEN, and J. Sunil RAO

We derive the best predictive estimator (BPE) of the fixed parameters under two well-known small area models, the Fay–Herriot model and the nested-error regression model. This leads to a new prediction procedure, called observed best prediction (OBP), which is different from the empirical best linear unbiased prediction (EBLUP). We show that BPE is more reasonable than the traditional estimators derived from estimation considerations, such as maximum likelihood (ML) and restricted maximum likelihood (REML), if the main interest is estimation of small area means, which is a mixed-model prediction problem. We use both theoretical derivations and empirical studies to demonstrate that the OBP can significantly outperform EBLUP in terms of the mean squared prediction error (MSPE), if the underlying model is misspecified. On the other hand, when the underlying model is correctly specified, the overall predictive performance of the OBP is very similar to that of the EBLUP if the number of small areas is large. A general theory about OBP, including its exact MSPE comparison with the BLUP in the context of mixed-model prediction, and asymptotic behavior of the BPE, is developed. A real data example is considered. A supplementary appendix is available online.

KEY WORDS: Fay–Herriot model; Mean squared prediction error (MSPE); Model misspecification; Nested-error regression model; Robustness.

1. INTRODUCTION

The empirical best linear unbiased prediction, or EBLUP, is well known in small area estimation (SAE, e.g., Rao 2003; Jiang and Lahiri 2006). There are several ways of deriving the BLUP (e.g., Jiang 2007, p. 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 $\boldsymbol{\beta}$ 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). However, in many cases, such as in SAE, the problem of main interest is prediction, rather than estimation. The implication is that the EBLUP may be regarded as a hybrid of optimal prediction (i.e., BP) and optimal estimation (e.g., ML). Nevertheless, if prediction is of main interest, it would be more natural to have a purely predictive procedure, in which both the predictor and estimator are derived from predictive considerations.

For example, the Fay–Herriot model (Fay and Herriot 1979) is widely used in SAE. It was proposed to estimate the per-capita income of small places with population size less than 1000. The model can be expressed in terms of a mixed effects model:

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

where $\mathbf{x}_i$ is a vector of known covariates, $\boldsymbol{\beta}$ 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 $v_i \sim \mathrm{N}(0, A)$ and $e_i \sim \mathrm{N}(0, D_i)$. 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 $\theta_i = \mathbf{x}_i'\boldsymbol{\beta} + v_i$, $1 \le i \le m$. Thus, our main interest is prediction of the mixed effects.

From a practical point of view, any proposed model is subject to model misspecification. In this article, we shall focus on misspecification of the mean functions. Similarly, a model corresponds to a specified form of the mean function. For example, for the Fay–Herriot model, the mean function is specified as $\mathbf{x}_i'\boldsymbol{\beta}$. However, it is (very) possible that the true underlying model is not the same as the assumed, or the assumed model is misspecified. We illustrate this with a real-data example. Morris and Christiansen (1995) presented a dataset involving 23 hospitals (out of a total of 219 hospitals) that had at least 50 kidney transplants during a 27-month period (see Table 4 in Section 7). The y_i's are graft failure rates for kidney transplant operations, that is, $y_i =$ number of graft failures$/n_i$, where n_i is the number of kidney transplants at hospital i during the period of interest. The variance for the graft failure rate, D_i, is approximated by $(0.2)(0.8)/n_i$, where 0.2 is the observed failure rate for all hospitals. Thus, D_i is assumed known. In addition, a severity index x_i is available for each hospital, which is the average fraction of females, blacks, children, and extremely ill kidney recipients at hospital i. Ganesh (2009) proposed a Fay–Herriot model for the graft failure rates, which is (1) with $\mathbf{x}_i'\boldsymbol{\beta} = \beta_0 + \beta_1 x_i$. Note that the graft failure rates are binomial proportion of fairly large denominators (at least 50). Thus, a normal distribution for the y_i's is not unreasonable, at least from an approximation point of view, by the central limit theorem. However, inspections of the raw data suggest some nonlinear trends in the mean function. See Figure 1 in the Appendix, Sections 4.2 and 7 for more details. This raises a concern about model misspecification.

Jiming Jiang is Professor, Department of Statistics, University of California, Davis, Davis, CA 95616 (E-mail: jiang@wald.ucdavis.edu). Thuan Nguyen is Assistant Professor, Department of Public Health and Preventive Medicine, Oregon Health and Science University, Portland, OR 97239-3098. J. Sunil Rao is Professor, Department of Epidemiology and Public Health, University of Miami, Miami, FL 33136. Jiming Jiang is partially supported by the NSF grant DMS-0809127. J. Sunil Rao is partially supported by the NSF grant DMS-0806076. The research of all three authors are partially supported by the NIH grant R01-GM085205A1. The authors are grateful to an Associate Editor and two Referees for their constructive comments that have led to improvements of the article.

Journal of the American Statistical Association
June 2011, Vol. 106, No. 494, Theory and Methods
DOI: 10.1198/jasa.2011.tm10221

On the other hand, the true small area means should not be dependent on the assumed model. Therefore, we need expressions of the mixed effects that do not depend on the assumed model. Note that $\theta_i = \mathrm{E}(y_i|v_i)$, $1 \le i \le m$. Now suppose that the true underlying model can be expressed as

$$y_i = \mu_i + v_i + e_i, \qquad i = 1, \ldots, m, \tag{2}$$

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, the bottom line is $\mathrm{E}(y_i) = \mu_i$, $1 \le i \le m$. Therefore, under model (2), the small area means can be expressed as

$$\theta_i = \mu_i + v_i = \mathrm{E}(y_i) + v_i, \qquad i = 1, \ldots, m. \tag{3}$$

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. The objective expression (3) of the mixed effects will be used later in deriving our best predictive estimator.

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

$$\mathrm{MSPE}(\tilde{\boldsymbol{\theta}}) = \mathrm{E}(|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}|^2) = \sum_{i=1}^{m} \mathrm{E}(\tilde{\theta}_i - \theta_i)^2. \tag{4}$$

See Section 8 for discussion on the definition of the MSPE. 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 *not* model dependent. This is a key to our approach. Under the MSPE measure, the BP of $\boldsymbol{\theta}$ is its conditional expectation, $\tilde{\boldsymbol{\theta}} = \mathrm{E}(\boldsymbol{\theta}|y)$. Under the assumed model (1), and given the parameters $\boldsymbol{\psi} = (\boldsymbol{\beta}', A)'$, the BP can be expressed as

$$\tilde{\boldsymbol{\theta}}(\boldsymbol{\psi}) = \mathrm{E}_{\mathrm{M},\boldsymbol{\psi}}(\boldsymbol{\theta}|y) = \left[\mathbf{x}_i'\boldsymbol{\beta} + \frac{A}{A+D_i}(y_i - \mathbf{x}_i'\boldsymbol{\beta})\right]_{1\le i\le m} \tag{5}$$

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

$$\begin{aligned}\mathrm{MSPE}\{\tilde{\boldsymbol{\theta}}(\boldsymbol{\beta})\} &= \sum_{i=1}^{m} \mathrm{E}\{B_i y_i - \theta_i + \mathbf{x}_i'\boldsymbol{\beta}(1 - B_i)\}^2 \\ &= I_1 + 2I_2 + I_3,\end{aligned} \tag{6}$$

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

$$\begin{aligned}&\mathrm{MSPE}\{\tilde{\boldsymbol{\theta}}(\boldsymbol{\beta})\} \\ &\quad = \mathrm{E}\left\{I_1 + \sum_{i=1}^{m}(1 - B_i)^2(\mathbf{x}_i'\boldsymbol{\beta})^2 - 2\sum_{i=1}^{m}(1 - B_i)^2\mathbf{x}_i'\boldsymbol{\beta} y_i\right\}.\end{aligned} \tag{7}$$

The right-hand side of (7) suggests a natural estimator of $\boldsymbol{\beta}$, by minimizing the expression inside the expectation, which is equivalent to minimizing

$$\begin{aligned}Q(\boldsymbol{\beta}) &= \sum_{i=1}^{m}(1 - B_i)^2(\mathbf{x}_i'\boldsymbol{\beta})^2 - 2\sum_{i=1}^{m}(1 - B_i)^2\mathbf{x}_i'\boldsymbol{\beta} y_i \\ &= \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{\Gamma}^2\mathbf{X}\boldsymbol{\beta} - 2\mathbf{y}'\boldsymbol{\Gamma}^2\mathbf{X}\boldsymbol{\beta},\end{aligned} \tag{8}$$

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

$$\begin{aligned}\tilde{\boldsymbol{\beta}} &= (\mathbf{X}'\boldsymbol{\Gamma}^2\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Gamma}^2\mathbf{y} \\ &= \left\{\sum_{i=1}^{m}(1 - B_i)^2\mathbf{x}_i\mathbf{x}_i'\right\}^{-1}\sum_{i=1}^{m}(1 - B_i)^2\mathbf{x}_i y_i.\end{aligned} \tag{9}$$

Here we assume, without loss of generality, that $\mathbf{X}$ is of full column rank. Note that $\tilde{\boldsymbol{\beta}}$ is different from the MLE of $\boldsymbol{\beta}$,

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\ &= \left(\sum_{i=1}^{m}\frac{\mathbf{x}_i\mathbf{x}_i'}{A + D_i}\right)^{-1}\sum_{i=1}^{m}\frac{\mathbf{x}_i y_i}{A + D_i},\end{aligned} \tag{10}$$

where $\mathbf{V} = \mathrm{diag}(A + D_i, 1 \le i \le m) = \mathrm{var}(\mathbf{y})$. While $\hat{\boldsymbol{\beta}}$ maximizes the likelihood function, $\tilde{\boldsymbol{\beta}}$ minimizes the "observed" MSPE which is the expression inside the expectation on the right-hand side of (7). We call $\tilde{\boldsymbol{\beta}}$ given by (9) the *best predictive estimator*, or BPE, of $\boldsymbol{\beta}$. Note that the BPE has the property that, theoretically, its expected value,

$$\mathrm{E}(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}'\boldsymbol{\Gamma}^2\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Gamma}^2\mathrm{E}(\mathbf{y}), \tag{11}$$

is the $\boldsymbol{\beta}$ that minimizes $\mathrm{MSPE}\{\tilde{\boldsymbol{\theta}}(\boldsymbol{\beta})\}$. However, the expression (11) is not computable. Another interesting observation is that the BPE (9) gives more weights to data points with larger sampling variances. Note that the BP is a weighted average of the direct estimator (i.e., y_i) and the "synthetic" estimator [i.e., $\mathbf{x}_i'\boldsymbol{\beta}$; see (5)], and it is the latter that is model dependent, and receives more weights for areas with larger sampling variances. Thus, the BPE strategy seems to make logical sense, because the model-based prediction method is more relevant to areas with larger sampling variances, therefore, the "voice" of those areas should be heard more (given more weights) in determining what parameters to use for the model-based prediction method. On the other hand, the MLE (10) does just the opposite—assigning more weights to data points with smaller sampling variances, which are less relevant to the model in terms of the prediction. Note that the assumed model is (heavily) used anyway, in the BP. The only difference is how to best estimate the parameters under the assumed model (BPE or MLE) in terms of the predictive performance. Finally, in the special case that all the D_i's are equal (the so-called balanced case), the BPE (9) and MLE (10) are identical.

A predictor of the mixed effects $\boldsymbol{\theta}$ is then obtained by replacing $\boldsymbol{\beta}$ in the BP (5) by its BPE (note that here A is assumed known). We call this predictor the *observed best predictor*, or OBP. The reason is that the BPE is the minimizer of the observed MSPE [see the note below (10)]. If the observed MSPE were the true MSPE, the BPE would give us the BP. However, because, instead, we are dealing with the observed MSPE, the corresponding predictor (obtained by the same procedure with the MSPE replaced by the observed MSPE) should be called the observed BP. Note that the OBP is different from the EBLUP, and this difference will be highlighted in subsequent sections.

The above derivation (for the special case) outlines the main idea of our approach, known as the best predictive SAE, or OBP. In the next section, we show by a simple example that the gain of OBP over EBLUP can be substantial, if the underlying model is misspecified. The BPE and OBP are derived for the general case of the Fay–Herriot model, in which A is unknown, and for another well-known class of small area models, the nested-error regression model (Battese, Harter, and Fuller 1988), in Section 3. In Section 4 we carry out a number of simulation studies that compare finite-sample performance of the OBP with several standard EBLUPs. The simulation studies cover different aspects of model misspecification as well as a case where the underlying model is correctly specified. The simple example of Section 2 and the simulation results of Section 4 suggest a general theory about the OBP, and this theory is developed in Section 5, including exact MSPE comparison of the OBP with the BLUP and asymptotic behavior of the BPE under model misspecifications. In Section 6 we consider estimation of the area-specific MSPE of the OBP, and propose a second-order unbiased MSPE estimator under possible model misspecification as well as a bootstrap MSPE estimator. The above real-data example is revisited in Section 7. Some discussion and concluding remarks are offered in Section 8. Technical results and further details are deferred to the Appendix that is supplementary to the article (available at *www.stat.ucdavis.edu/~jiang/jnr.suppl.pdf*).

2. A SIMPLE EXAMPLE

In this section, we use a very simple example to show that the gain of OBP over BLUP or EBLUP can be substantial, if the underlying model is misspecified. Consider a special case of the Fay–Herriot model (1), in which $\mathbf{x}_i'\boldsymbol{\beta} = \beta$, an unknown mean. To make it even simpler, suppose that A is known, so that one can actually compute the BLUP. Furthermore, suppose that $m = 2n$, $D_i = a, 1 \le i \le n$, and $D_i = b, n+1 \le i \le m$, where a, b are positive known constants.

Now suppose that, actually, the underlying model is

$$\begin{aligned} y_i &= c + v_i + e_i, \qquad 1 \le i \le n, \qquad \text{and} \\ y_i &= d + v_i + e_i, \qquad n+1 \le i \le m, \end{aligned} \tag{12}$$

where $c \ne d$; in other words, we have a model misspecification by assuming $c = d$.

For this special case, we can actually derive the exact expressions of the MSPEs. It can be shown (see the Appendix) that for the OBP $\tilde{\boldsymbol{\theta}}$, we have

$$\begin{aligned} \mathrm{MSPE}(\tilde{\boldsymbol{\theta}}) &= \left\{\left(\frac{a}{A+a} + \frac{b}{A+b}\right)A \right. \\ &\quad \left. + \frac{a^2b^2(c-d)^2}{a^2(A+b)^2 + b^2(A+a)^2}\right\}n \\ &\quad + \frac{a^4(A+b)^3 + b^4(A+a)^3}{(A+a)(A+b)\{a^2(A+b)^2 + b^2(A+a)^2\}} \\ &= g_1 n + h_1, \end{aligned} \tag{13}$$

while for the BLUP $\hat{\boldsymbol{\theta}}$, we have

$$\begin{aligned} \mathrm{MSPE}(\hat{\boldsymbol{\theta}}) &= \left\{\left(\frac{a}{A+a} + \frac{b}{A+b}\right)A + \frac{(a^2+b^2)(c-d)^2}{(2A+a+b)^2}\right\}n \\ &\quad + \frac{a^2(A+b)^2 + b^2(A+a)^2}{(A+a)(A+b)(2A+a+b)} \\ &= g_2 n + h_2. \end{aligned} \tag{14}$$

If $c \ne d$, then $g_1 \le g_2$ and $h_1 \ge h_2$ with equality holding in both cases if and only if

$$a^2(A+b) = b^2(A+a). \tag{15}$$

Now suppose that (15) does not hold, then we have $g_1 < g_2$, and $h_1 > h_2$, but the latter is not important when n is large. In fact, we have

$$\lim_{n\to\infty}\{\mathrm{MSPE}(\tilde{\boldsymbol{\theta}})/\mathrm{MSPE}(\hat{\boldsymbol{\theta}})\} = g_1/g_2 < 1. \tag{16}$$

For example, suppose that $A/(c-d)^2 \approx 0$, $A/b \approx 0$, and $b/a \approx 0$. Then it is easy to show that $g_1/g_2 \approx 0.5$. Thus, in this case, the MSPE of the OBP is asymptotically about half of that of the BLUP. On the other hand, if $c = d$, that is, if the underlying model is correctly specified, then we have $g_1 = g_2$ while, still, $h_1 \ge h_2$. Therefore, in this case, $\mathrm{MSPE}(\tilde{\boldsymbol{\theta}}) \ge \mathrm{MSPE}(\hat{\boldsymbol{\theta}})$. However, we have $\lim_{n\to\infty}\mathrm{MSPE}(\tilde{\boldsymbol{\theta}})/\mathrm{MSPE}(\hat{\boldsymbol{\theta}}) = 1$, hence, in this case, the MSPEs of the OBP and BLUP are asymptotically the same.

An extension of this story, in much more generality, will be given in Section 5.1.

3. OBP FOR TWO IMPORTANT CLASSES OF SAE MODELS

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-hand 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, $\tilde{\boldsymbol{\theta}}(\boldsymbol{\psi}) = \mathbf{y} - \boldsymbol{\Gamma}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, where $\boldsymbol{\Gamma}$ is defined below (8). By (2) and (3), it can be shown that

$$\begin{aligned} &\mathrm{MSPE}\{\tilde{\boldsymbol{\theta}}(\boldsymbol{\psi})\} \\ &\quad = \mathrm{E}\{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Gamma}^2(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + 2A\,\mathrm{tr}(\boldsymbol{\Gamma}) - \mathrm{tr}(\mathbf{D})\}, \end{aligned} \tag{17}$$

where $\mathbf{D} = \mathrm{diag}(D_i, 1 \le i \le m)$. The BPE of $\boldsymbol{\psi} = (\boldsymbol{\beta}', A)'$ is obtained by minimizing the expression inside the E on the right-hand side of (17), which is equivalent to minimizing

$$Q(\boldsymbol{\psi}) = (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Gamma}^2(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + 2A\,\mathrm{tr}(\boldsymbol{\Gamma}). \tag{18}$$

Let $\tilde{Q}(A)$ be $Q(\boldsymbol{\psi})$ with $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$ given by (9). It can be shown that $\tilde{Q}(A) = \mathbf{y}'\boldsymbol{\Gamma} P_{(\boldsymbol{\Gamma}\mathbf{X})^\perp}\boldsymbol{\Gamma}\mathbf{y} + 2A\,\mathrm{tr}(\boldsymbol{\Gamma})$, 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, $P_{(\boldsymbol{\Gamma}\mathbf{X})^\perp} = \mathbf{I}_m - \boldsymbol{\Gamma}\mathbf{X}(\mathbf{X}'\boldsymbol{\Gamma}^2\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Gamma}$ 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 $\boldsymbol{\beta}$ is given by (9) with A replaced by $\tilde{A}$. Given the BPE of $\boldsymbol{\psi}$, $\tilde{\boldsymbol{\psi}} = (\tilde{\boldsymbol{\beta}}', \tilde{A})'$, the OBP of $\boldsymbol{\theta}$ is given by the BP (5) with $\boldsymbol{\psi} = \tilde{\boldsymbol{\psi}}$.

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 superpopulation nested-error regression model (Battese, Harter, and Fuller 1988) holds:

$$Y_{ik} = \mathbf{X}_{ik}'\boldsymbol{\beta} + v_i + e_{ik}, \qquad i = 1, \ldots, m, k = 1, \ldots, N_i, \quad (19)$$

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 $v_i \sim \mathrm{N}(0, \sigma_v^2)$ and $e_{ik} \sim \mathrm{N}(0, \sigma_e^2)$. 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 $\boldsymbol{\psi} = (\boldsymbol{\beta}', \sigma_v^2, \sigma_e^2)'$ denote the vector of parameters under the nested-error regression model (19). Under this model with $\boldsymbol{\psi}$ being the true parameter vector, the BP for μ_i is $\mathrm{E}_{\mathrm{M},\boldsymbol{\psi}}(\mu_i|y) = N_i^{-1}\{\sum_{j=1}^{n_i} y_{ij} + \sum_{k \notin I_i}\mathrm{E}_{\mathrm{M},\boldsymbol{\psi}}(Y_{ik}|y_i)\}$, which can be expressed as

$$\tilde{\mu}_i(\boldsymbol{\psi}) = \bar{\mathbf{X}}_i'\boldsymbol{\beta} + \left\{\frac{n_i}{N_i} + \left(1 - \frac{n_i}{N_i}\right)\frac{n_i\sigma_v^2}{\sigma_e^2 + n_i\sigma_v^2}\right\}(\bar{y}_{i\cdot} - \bar{\mathbf{x}}_{i\cdot}'\boldsymbol{\beta}), \quad (20)$$

where $\mathrm{E}_{\mathrm{M},\boldsymbol{\psi}}$ denotes the model-based conditional expectation given that $\boldsymbol{\psi}$ 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 (20) 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. This is because the latter is almost free of model assumptions, and therefore robust to model misspecifications. Note that the situation is different under the Fay–Herriot model, because the sampling data are not available at the unit level, therefore, it is not possible to evaluate the design-based MSPE. Instead, we considered a model with very weak assumptions for the Fay–Herriot model, that is, (2). As discussed in Section 8 (first paragraph), the idea of OBP can be viewed broadly as entertaining two models, a broader model and a more restrictive one. Although there is no strict rule on how to choose the broader model, under which the MSPE is evaluated, a guideline is that the assumption about the broader model should be as weak as possible, as long as the MSPE can be expressed as the expectation of a function of the observed data and the parameters involved in the BP [see (7), (17), and (22) below]. The design-based MSPE is given by

$$\begin{aligned}\mathrm{MSPE}\{\tilde{\boldsymbol{\mu}}(\boldsymbol{\psi})\} &= \mathrm{E}_{\mathrm{d}}\{|\tilde{\boldsymbol{\mu}}(\boldsymbol{\psi}) - \boldsymbol{\mu}|^2\} \\ &= \sum_{i=1}^{m}\mathrm{E}_{\mathrm{d}}\{\tilde{\mu}_i(\boldsymbol{\psi}) - \mu_i\}^2, \qquad (21)\end{aligned}$$

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

$$\mathrm{MSPE}\{\tilde{\boldsymbol{\mu}}(\boldsymbol{\psi})\} = \mathrm{E}_{\mathrm{d}}\left[\sum_{i=1}^{m}\{\tilde{\mu}_i^2(\boldsymbol{\psi}) - 2a_i(\sigma_v^2, \sigma_e^2)\bar{\mathbf{X}}_i'\boldsymbol{\beta}\bar{y}_{i\cdot} + b_i(\sigma_v^2, \sigma_e^2)\hat{\mu}_i^2\}\right], \quad (22)$$

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

$$\hat{\mu}_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{N_i - 1}{N_i(n_i - 1)}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2 \quad (23)$$

(see the Appendix). Thus, the BPE of $\boldsymbol{\psi}$ is obtained by minimizing

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

A computational procedure for the BPE similar to that for the Fay–Herriot model can be given (see the Appendix). Given the BPE $\tilde{\boldsymbol{\psi}} = (\tilde{\boldsymbol{\beta}}', \tilde{\sigma}_v^2, \tilde{\sigma}_e^2)'$, the OBP of μ_i is given by $\tilde{\mu}_i = \tilde{\mu}_i(\tilde{\boldsymbol{\psi}})$, $1 \leq i \leq m$, where $\tilde{\mu}_i(\boldsymbol{\psi})$ is given by (20).

4. SIMULATION STUDIES

4.1 A Simple Example

We first use a simple, simulated example to demonstrate the theoretical properties of the OBP and its comparison to the EBLUP (see Section 5 for details). The example is the same as the simple Fay–Herriot model considered in Section 2 except that now A is unknown. Note that the assumed model can be written as $y_i = \beta + v_i + e_i, i = 1, \ldots, m$, where the v_i's and e_i's are independent such that $v_i \sim \mathrm{N}(0, A)$ with A unknown, and $e_i \sim \mathrm{N}(0, D_i)$ with D_i given in Section 2 and further specified below. We consider four different estimators of A that have been traditionally used. These are the ML, REML, Fay–Herriot (F–H; Fay and Herriot 1979; also see Pfeffermann and Nathan 1981 and Datta, Rao, and Smith 2005) and Prasad–Rao (P–R; Prasad and Rao 1990) estimators. Given one of the estimators of A, denoted by $\hat{A}$, the EBLUP is obtained by the following steps: (i) computing $\hat{\beta}$ of (10) with $x_i = 1$ and A replaced by $\hat{A}$;

and (ii) computing the EBLUP $\hat{\boldsymbol{\theta}} = (\hat{\theta}_i)_{1\le i\le m}$ of $\boldsymbol{\theta} = (\theta_i)_{1\le i\le m}$, where

$$\begin{aligned}\hat{\theta}_i &= \frac{\hat{A}}{\hat{A}+a}y_i + \frac{a}{\hat{A}+a}\hat{\beta}, \qquad 1\le i\le n;\\ \hat{\theta}_i &= \frac{\hat{A}}{\hat{A}+b}y_i + \frac{b}{\hat{A}+b}\hat{\beta}, \qquad n+1\le i\le m\end{aligned} \tag{24}$$

[note that the $\hat{A}$ in (24) is the same $\hat{A}$ used to compute $\hat{\beta}$]. Depending on whether $\hat{A}$ is the ML, REML, F–H, or P–R estimator, the corresponding EBLUPs are denoted by EBLUP-1, EBLUP-2, EBLUP-3, and EBLUP-4, respectively. The OBP of $\boldsymbol{\theta}$, $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_i)_{1\le i\le m}$, is given by the right-hand sides of (24) with $\hat{A}$ and $\hat{\beta}$ replaced by $\tilde{A}$ and $\tilde{\beta}$, respectively, where $\tilde{\boldsymbol{\psi}} = (\tilde{\beta}', \tilde{A})'$ is the BPE of $\boldsymbol{\psi} = (\beta', A)'$.

The values of a, b, c, d are chosen as $a = 4$, $b = 1$, $c = 0$, while $d = 1$ or 5. The simulation is run for three different sample sizes: $m = 50$, 100, and 200. In each case, $K = 500$ simulation runs are carried out. Within each simulation run, a dataset is generated under the true underlying model (12) with $A = 0.2$ (thus, we are in a situation where the underlying model is misspecified). Then, five predictors, EBLUP-1, EBLUP-2, EBLUP-3, EBLUP-4, and OBP, are computed based on the same generated dataset. Let $\check{\boldsymbol{\theta}}^{(k)}$ denote one of these five predictors of $\boldsymbol{\theta}$ based on the dataset generated in the kth simulation run, and $\boldsymbol{\theta}^{(k)}$ be the true $\boldsymbol{\theta}$ for the same simulation run, $k = 1, \ldots, K$. Note that the true $\boldsymbol{\theta} = (\theta_i)_{1\le i\le m}$ is given by $\theta_i = c + v_i$, $1 \le i \le n$, and $\theta_i = d + v_i$, $n + 1 \le i \le m$. The simulated, or empirical MSPE for a particular kind of predictor is then given by $\text{MSPE}^* = K^{-1}\sum_{k=1}^K |\check{\boldsymbol{\theta}}^{(k)} - \boldsymbol{\theta}^{(k)}|^2 = K^{-1}\sum_{k=1}^K\sum_{i=1}^m \{\check{\theta}_i^{(k)} - \theta_i^{(k)}\}^2$.

The empirical MSPEs are reported in Table 1. The number in the parentheses is the percentage increase in MSPE by the corresponding EBLUP over the OBP. The Monte Carlo standard errors (SEs) are reported in Table 1 of the Appendix. It is seen that all the EBLUPs perform very similarly, while the OBP is a distance away from the EBLUPs. The MSPEs of the EBLUPs are somewhere between 11% to 44% higher than that of the OBP. The percentage increase in MSPE by the EBLUPs is more substantial for $d = 5$ than for $d = 1$. This makes sense because $d = 5$ features a more serious model misspecification than $d = 1$ does, and the OBP shines when the underlying model is misspecified. A more explicit explanation of this pattern may be seen from the comparison of the formulae for the exact MSPEs, that is, (13) and (14) (with $c = 0$), although the latter are derived for the case that A is known. The formulae show that the difference in MSPE between the OBP and any of the EBLUPs gets larger, even proportionally, as d increases.

We next compare the OBP and EBLUPs in terms of area-specific MSPEs. Although the OBP is defined by minimizing the overall (observed) MSPE, there is no guarantee that its area-specific MSPEs are minimal. On the other hand, area-specific MSPEs are often of main interest in SAE. Therefore, a comparison of the area-specific MSPEs of the OBP with those of the EBLUP is important, especially from a practical point of view. Such a comparison may also provide further details for the overal MSPEs reported in Table 1. The empirical area-specific MSPE is evaluated by $\text{MSPE}_i^* = K^{-1}\sum_{k=1}^K\{\check{\theta}_i^{(k)} - \theta_i^{(k)}\}^2$ with $K = 500$, where $\check{\theta}_i^{(k)}$ and $\theta_i^{(k)}$ are the predictor and true small area mean, respectively, for the ith small area in the kth simulation run, $1 \le i \le m$, $1 \le k \le K$. Due to the fairly large number of small areas involved ($m = 50$, 100, or 200 in our simulations), we summarize the results using boxplots and histograms, as shown in Figures 1 and 2. The figures reveal some untold stories by the overall MSPEs. First, the boxplots show a significant difference in the distributions of the empirical MSPEs between the OBP and EBLUPs. Not only does the OBP have smaller median empirical MSPE in each case, what is more apparent is the range, or variation, of the empirical MSPEs overwhelmingly in favor of the OBP. Second, the histograms exhibit quite different shapes between the OBP and EBLUPs. A closer look at the numbers shows that the empirical MSPEs of the EBLUPs are somewhere between slightly to moderately smaller than those of the OBP for half of the small areas; but for the other half of the small areas, the empirical MSPEs of the EBLUPs are much larger than those of the OBP. The pattern can also be seen from the histograms. Recall the assumed model has a common mean for all the small areas, while the true model has one mean for half of the areas and another mean for the other half. Apparently, what the EBLUP does is to "side with" one mean while "abandoning" the other. The OBP, on the other hand, uses a rather different strategy, by "staying in the middle" or "balancing" between the two means. This explains the bimodal histograms for the EBLUPs, compared to the fairly normal-looklike histograms for the OBP (with much narrower spreads). Overall, the simulation results show a much more robust performance of the OBP in terms of the area-specific MSPE as compared to the EBLUPs.

4.2 An Example Imitating the Hospital Data

We now consider an example that, in a way, imitates the hospital data (or at least the models proposed for the data). Unlike the previous example, here we consider the case where there is a "slight" misspecification of the underlying model. Recall the hospital data discussed in Section 1. An inspection of the scatter plot (see Figure 1 in the Appendix) suggests that a quadratic

Table 1. Empirical MSPE (% increase over OBP)

m	d	EBLUP-1	EBLUP-2	EBLUP-3	EBLUP-4	OBP
50	1	28.76 (28%)	27.94 (25%)	25.00 (11%)	25.87 (15%)	22.43
100		51.05 (26%)	50.20 (24%)	47.74 (18%)	49.02 (21%)	40.42
200		94.22 (26%)	93.95 (26%)	92.52 (24%)	93.87 (25%)	74.86
50	5	95.83 (42%)	95.05 (41%)	93.55 (39%)	93.12 (38%)	67.41
100		189.93 (44%)	189.22 (43%)	186.51 (41%)	185.61 (41%)	132.01
200		372.59 (44%)	371.92 (44%)	366.96 (42%)	365.21 (41%)	258.60

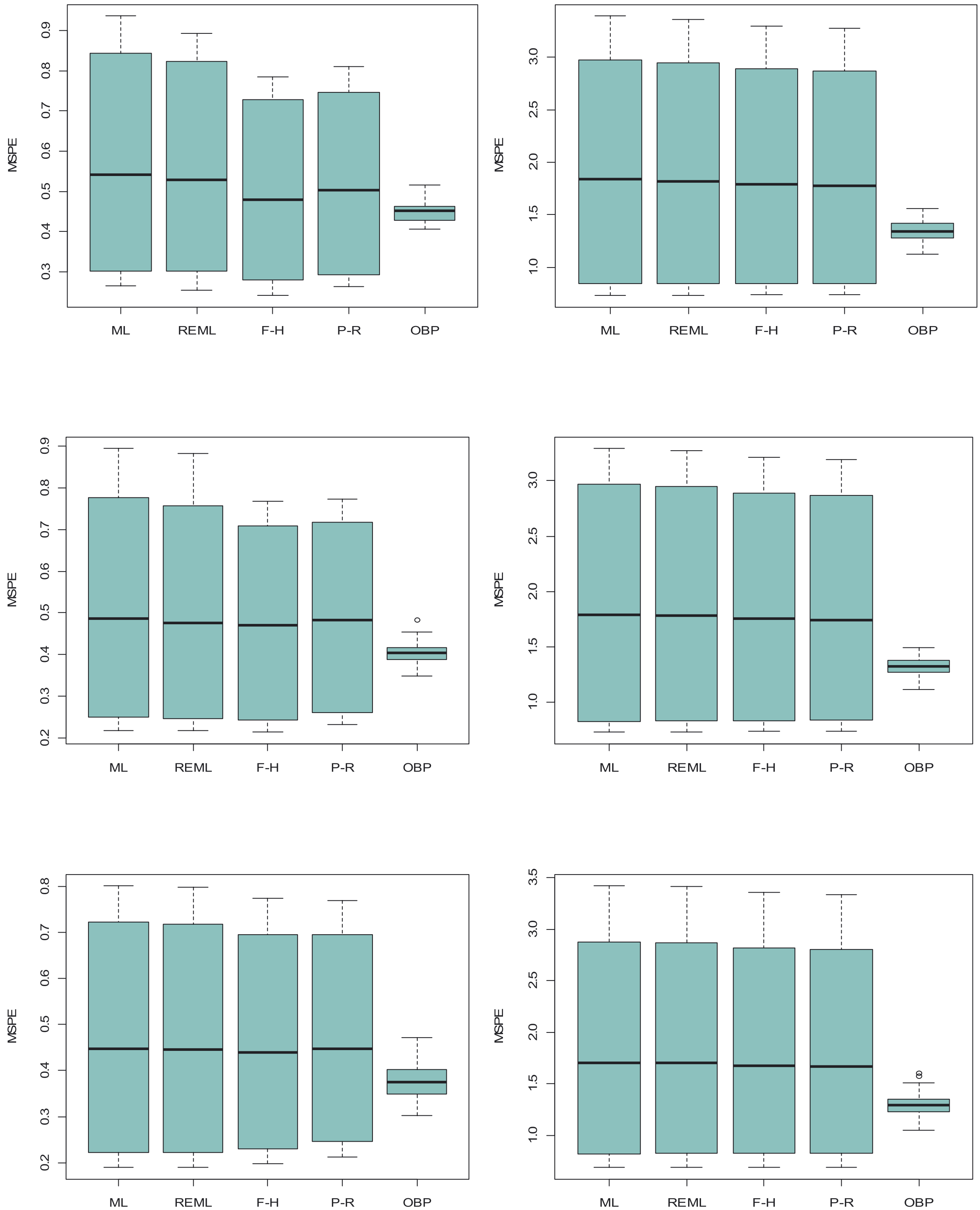

Figure 1. Boxplots of $\text{MSPE}_i^*, i = 1, \ldots, m$. Upper left: $m = 50, d = 1$; Upper right: $m = 50, d = 5$; Middle left: $m = 100, d = 1$; Middle right: $m = 100, d = 5$; Lower left: $m = 200, d = 1$; Lower right: $m = 200, d = 5$. Within each plot from left to right: EBLUP-1, EBLUP-2, EBLUP-3, EBLUP-4, and OBP. The online version of this figure is in color.

model would fit the data well except for the point at the upper right corner. Jiang, Nguyen, and Rao (2010) proposed a cubic model (see the smooth curve in the figure) for the same data. On the other hand, there has been a concern that the point at the upper right corner might be an "outlier," in some way, so a cubic model to accommodate a single point might overfit the data. We set up the simulation to investigate this outlying effect. The data is generated from a quadratic-outlying (Q–O) model, as follows. Suppose that there is an abrupt "jump" in the mean response when x_i is greater than 0.3; otherwise, the mean re-

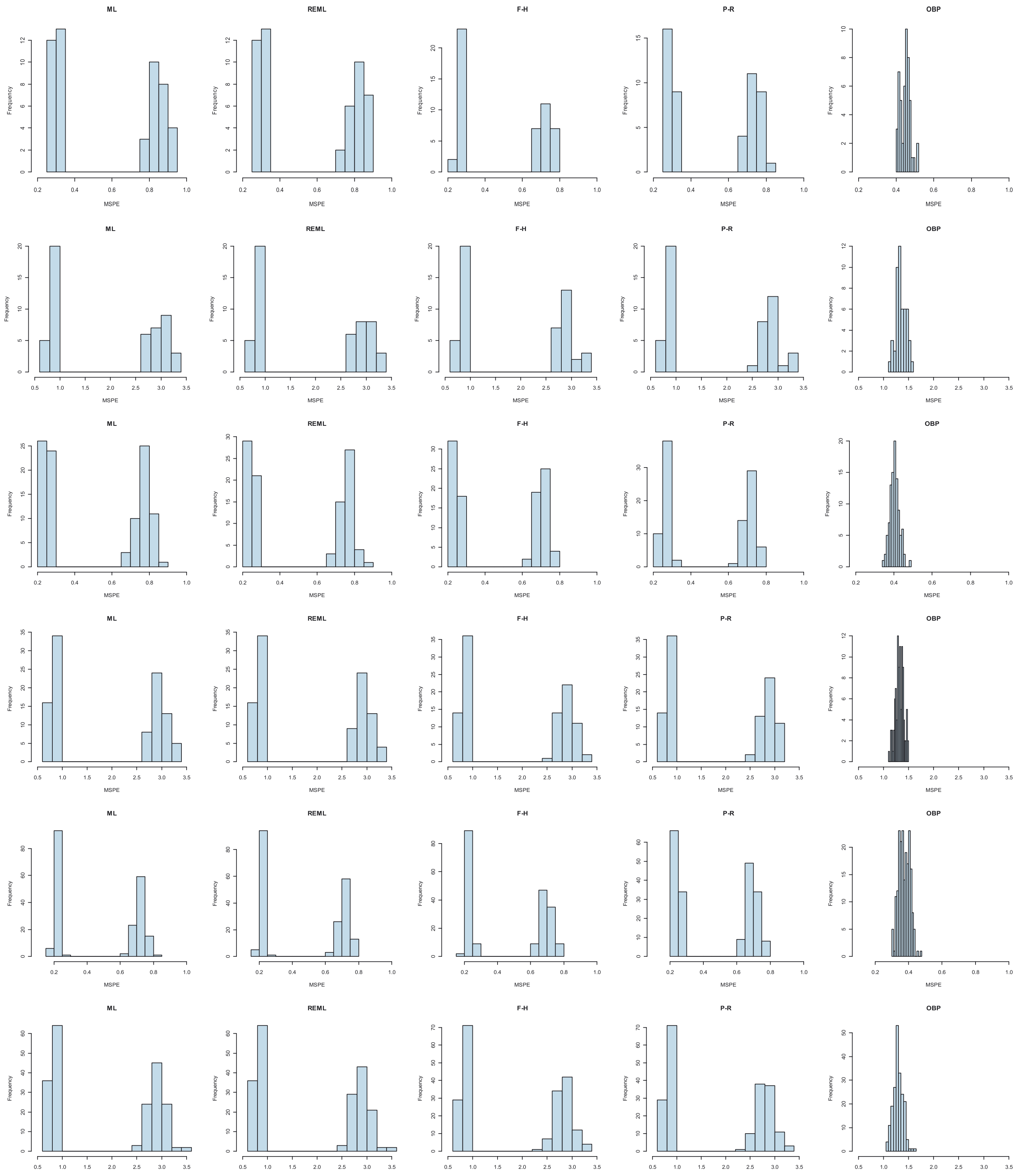

Figure 2. Histograms of MSPE_i^*, $i = 1, \ldots, m$. The rows, from top to bottom, correspond to $m = 50, d = 1$; $m = 50, d = 5$; $m = 100, d = 1$; $m = 100, d = 5$; $m = 200, d = 1$; and $m = 200, d = 5$, respectively. Within each row from left to right: EBLUP-1, EBLUP-2, EBLUP-3, EBLUP-4, and OBP. Within each row the x-axes have the same range. The online version of this figure is in color.

sponse is a quadratic function of the covariate. This can be expressed as

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + d1_{(x_i > 0.3)} + v_i + e_i, \qquad (25)$$

$i = 1, \ldots, m$, where the x_i's are the same as the severity index of the hospital data, given in Table 4 of Section 7; the v_i's are iid $\mathrm{N}(0, A)$; the e_i's are independent with $e_i \sim \mathrm{N}(0, D_i)$, and the

D_i's are the same as in the hospital data (see Table 4 of Section 7). The true regression coefficients are $\beta_0 = -1.1$, $\beta_1 = 20$, $\beta_2 = -50$, and $d = 0.9$. These coefficents are chosen so that the quadratic function is concave (same as for the hospital data), the range of the response y_i is confined between 0 and 1 most of the time, and the height of the outlier (corresponding to the largest x_i) is about the same as the highest point of the nonoutliers, hence, in a way, mimic the hospital data. The true value of A is chosen as 0.0016, which is approximately equal to the average of the D_i's.

We consider two situations in our simulation. In the first case (Case I), the model is "slightly" misspecified; namely, the assumed model is a quadratic model, which is (25) without the term $d1_{(x_i>0.3)}$. (See a remark in the sequel on a situation of greater model misspecification.) In the second case (Case II), the model is correctly specified, that is, the Q–O model. As mentioned, a number of methods are available in estimating A. These are all consistent estimators, if the model is correctly specified. In case of the model misspecification, the estimators are not expected to be consistent, but it is reasonable to require that they are not too far from the true parameter. The reason is that we are still counting on that the assumed model is correct (otherwise, why would the prediction be based on the BP, which is derived under the assumed model?). On the other hand, we would also like to do well predictively in case that the model is slightly off. If A is known, it is generally agreed that one should use the true A in computing the predictor, because then the latter would be closest to the BP when the model is correct. Thus, if the model is slightly misspecified, which is probably the more realistic situation in practice, and A is unknown, one would prefer that the estimator of A still be reasonably close to the true value (note that here we only consider misspecification of the mean, so there is always a true value of A). In case there is a continuous covariate, such as the current case, it is shown in the Appendix that the following is an asymptotic upper bound for A:

$$U = (1+\delta)\left\{\frac{1}{m}\sum_{i=1}^{m}(y_i - \mathbf{X}_i'\hat{\boldsymbol{\beta}})^2 - \bar{D}\right\}, \tag{26}$$

where δ is an arbitrary positive constant, $\mathbf{X}_i$ is any vector that involves the covariates, and $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y$ with $\mathbf{X} = (\mathbf{X}_i')_{1\le i\le m}$. Furthermore, the upper bound holds under the general Fay–Herriot model (2), and thus is not affected by the model misspecification. The idea is to choose $\mathbf{X}_i$ that slightly expands the current vector of covariates (not to expand too much to avoid overfitting). In the current case, we let $\mathbf{X}_i$ include the next higher power of x_i. Also, the positive constant δ is chosen as 0.1. On the other hand, an obvious lower bound for A is 0. An estimator of A is truncated if it falls outside $[0, U]$.

We consider three different sample sizes: $m = 23$, $m = 115$, and $m = 230$. The first case is exactly as for the hospital data. In the second and third cases, we duplicate the design (i.e., the x_i's and D_i's) 5 times and 10 times, respectively. Before we report the results, let us first speculate on what to expect. There is a general theory about OBP. The theory is first derived for a special case in Section 2, then hinted by the simulation results of Section 4.1, and later further developed in Section 5. The following is what the theory would predict: (i) In Case I, the OBP has smaller overall MSPE compared to the EBLUPs; the overall advantage of the OBP over the EBLUPs will persist and stabilize as m increases. (ii) In Case II, the EBLUPs have smaller overall MSPE compared to the OBP; the overall advantage of the EBLUPs over the OBP will vanish as m increases. Table 2 reports the empirical overall MSPE results for EBLUP-1, EBLUP-4, and OBP (the results for the other EBLUPs are very similar). The results for $m = 115$ and $m = 230$ are based on $K = 500$ simulations; the results for $m = 23$ are based on $K = 5000$ simulations. As before, the number in the parentheses is the percentage increase in MSPE over the OBP, with negative percentage indicating decrease. The Monte Carlo SEs are reported in Table 2 of the Appendix. The simulation results show a trend that is inline with what the theory predicts.

Table 2. Empirical MSPE (% increase over OBP)

m	Case	EBLUP-1	EBLUP-4	OBP
23	I	0.04205 (1.42%)	0.04194 (1.16%)	0.04146
23	II	0.02519 (−1.13%)	0.02482 (−2.56%)	0.02547
115	I	0.14872 (2.12%)	0.14855 (2.00%)	0.14563
115	II	0.09459 (−0.99%)	0.09447 (−1.12%)	0.09554
230	I	0.28629 (2.30%)	0.28602 (2.20%)	0.27984
230	II	0.18457 (−0.69%)	0.18468 (−0.63%)	0.18585

As for the area-specific MSPEs, we report the percentage of small areas for which the empirical area-specific MSPEs of the OBP are smaller than those of the EBLUP-1 and EBLUP-4, respectively, as follows: For $m = 23$, 52% for both EBLUP-1 and EBLUP-4 under Case I; 57% for EBLUP-1 and 22% for EBLUP-4 under Case II; for $m = 115$, 56% for EBLUP-1 and 55% for EBLUP-4 under Case I; 17% for EBLUP-1 and 10% for EBLUP-4 under Case II; for $m = 230$, 57% for EBLUP-1 and 56% for EBLUP-4 under Case I; 20% for EBLUP-1 and 16% for EBLUP-4 under Case II. Note that, in Case I, OBP not only does better in most of the small areas, it does considerably or much better in a few small areas, including the outlying one, which is how its overall advantage is built. On the other hand, in Case II, despite the fact that the EBLUPs do better in by far most of the small areas (except for one case, in which the OBP actually does better in most of the small areas), the overall advantage of the EBLUP is vanishing when m increases, as shown in Table 2. Such features have been suggested by the simple example of Section 2, and will be further studied theoretically in the next section. In addition, estimation of the area-specific MSPEs of the OBP will be considered in Section 6.

Recall that in this subsection, we focus on the case where there is a slight (or no) misspecification of the underlying model. If we were to consider cases of more severe model misspecifications, the gains by OBP would of course be bigger. For example, if the assumed model is a linear model, that is, the right-hand side of (25) with only the first two terms, plus $v_i + e_i$, then for $m = 23$, the (overall) MSPEs of the EBLUPs are about 9% higher than that of the OBP, based on our simulation results (not shown).

5. A GENERAL THEORY ABOUT OBP

5.1 Exact MSPE Comparison With BLUP

This subsection may be regarded as an extended version of Section 2. Consider a general mixed-model prediction problem (e.g., Robinson 1991). The assumed model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e}, \tag{27}$$

where $\mathbf{X}, \mathbf{Z}$ are known matrices; $\boldsymbol{\beta}$ is a vector of fixed effects; $\mathbf{v}, \mathbf{e}$ are vectors of random effects and errors, respectively, such that $\mathbf{v} \sim \mathrm{N}(0, \mathbf{G})$, $\mathbf{e} \sim \mathrm{N}(0, \boldsymbol{\Sigma})$, and $\mathbf{v}, \mathbf{e}$ are uncorrelated. Suppose that the true underlying model is

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Z}\mathbf{v} + \mathbf{e}, \tag{28}$$

where $\boldsymbol{\mu} = \mathrm{E}(\mathbf{y})$. Here again, E without a subscript represents expectation with respect to the true distribution, which may be unknown but is not model dependent. Our interest is prediction of a vector of mixed effects that can be expressed as

$$\boldsymbol{\theta} = \mathbf{F}'\boldsymbol{\mu} + \mathbf{R}'\mathbf{v}, \tag{29}$$

where $\mathbf{F}, \mathbf{R}$ are known matrices. For example, for the Fay–Herriot model considered in the previous sections, we have $\mathbf{F} = \mathbf{R} = \mathbf{I}_m$, hence $\boldsymbol{\theta} = \boldsymbol{\mu} + \mathbf{v}$. Suppose that $\mathbf{G}, \boldsymbol{\Sigma}$ are known. Then, the BP of $\boldsymbol{\theta}$, in the sense of minimum MSPE, under the assumed model is given by $\mathrm{E}_{\mathrm{M}}(\boldsymbol{\theta}|\mathbf{y}) = \mathbf{F}'\boldsymbol{\mu} + \mathbf{R}'\mathrm{E}_{\mathrm{M}}(\mathbf{v}|\mathbf{y}) = \mathbf{F}'\mathbf{X}\boldsymbol{\beta} + \mathbf{R}'\mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, where E_{M} denotes expectation under the assumed model, $\mathbf{V} = \mathrm{var}(\mathbf{y}) = \boldsymbol{\Sigma} + \mathbf{Z}\mathbf{G}\mathbf{Z}'$, and $\boldsymbol{\beta}$ is the true vector of fixed effects, under the assumed model. If we write $\mathbf{B} = \mathbf{R}'\mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}$ and $\boldsymbol{\Gamma} = \mathbf{F}' - \mathbf{B}$, then the BP can be expressed in a slightly different way as

$$\mathrm{E}_{\mathrm{M}}(\boldsymbol{\theta}|\mathbf{y}) = \mathbf{F}'\mathbf{y} - \boldsymbol{\Gamma}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \tag{30}$$

It is easy to verify that the notation $\mathbf{B}$ and $\boldsymbol{\Gamma}$ are consistent with those used in Sections 1 and 3 under the Fay–Herriot model. Now let $\breve{\boldsymbol{\theta}}$ denote the right-hand side of (30) with a fixed, but arbitrary $\boldsymbol{\beta}$. Then, by (28) and (29), we have $\breve{\boldsymbol{\theta}} - \boldsymbol{\theta} = \mathbf{H}'\mathbf{v} + \mathbf{F}'\mathbf{e} - \boldsymbol{\Gamma}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, where $\mathbf{H} = \mathbf{Z}'\mathbf{F} - \mathbf{R}$. Thus, we have $\mathrm{MSPE}(\breve{\boldsymbol{\theta}}) = \mathrm{E}(|\breve{\boldsymbol{\theta}} - \boldsymbol{\theta}|^2) = \mathrm{E}(|\mathbf{H}'\mathbf{v} + \mathbf{F}'\mathbf{e}|^2) - 2\mathrm{E}\{(\mathbf{v}'\mathbf{H} + \mathbf{e}'\mathbf{F})\boldsymbol{\Gamma}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\} + \mathrm{E}\{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Gamma}'\boldsymbol{\Gamma}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\} = I_1 - 2I_2 + I_3$. Note that the I's here are different from those in (6). It is easy to see that I_1 does not depend on $\boldsymbol{\beta}$. Furthermore, I_2 does not depend on $\boldsymbol{\beta}$ either, because $I_2 = \mathrm{E}\{(\mathbf{v}'\mathbf{H} + \mathbf{e}'\mathbf{F})\boldsymbol{\Gamma}(\mathbf{y} - \boldsymbol{\mu})\} + \mathrm{E}\{(\mathbf{v}'\mathbf{H} + \mathbf{e}'\mathbf{F})\boldsymbol{\Gamma}(\boldsymbol{\mu} - \mathbf{X}\boldsymbol{\beta})\} = \mathrm{E}\{(\mathbf{v}'\mathbf{H} + \mathbf{e}'\mathbf{F})\boldsymbol{\Gamma}(\mathbf{Z}\mathbf{v} + \mathbf{e})$, by (28). Thus, we express the MSPE as

$$\mathrm{MSPE}(\breve{\boldsymbol{\theta}}) = \mathrm{E}\{I_1 - 2I_2 + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Gamma}'\boldsymbol{\Gamma}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\}. \tag{31}$$

The BPE of $\boldsymbol{\beta}$ is obtained by minimizing the expression inside the expectation, that is, $Q(\boldsymbol{\beta}) = I_1 - 2I_2 + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Gamma}'\boldsymbol{\Gamma}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$. Thus, the BPE is given by

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\mathbf{y}, \tag{32}$$

assuming that $\boldsymbol{\Gamma}'\boldsymbol{\Gamma}$ is nonsingular and $\mathbf{X}$ is full rank. Once the BPE is obtained, the OBP of $\boldsymbol{\theta}$ is given by the right-hand side of (30) with $\boldsymbol{\beta}$ replaced by $\tilde{\boldsymbol{\beta}}$, that is,

$$\tilde{\boldsymbol{\theta}} = \mathbf{F}'\mathbf{y} - \boldsymbol{\Gamma}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}). \tag{33}$$

On the other hand, the MLE of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \tag{34}$$

Thus, the BLUP of $\boldsymbol{\theta}$ is given by the right-hand side of (30) with $\boldsymbol{\beta}$ replaced by $\hat{\boldsymbol{\beta}}$, that is,

$$\hat{\boldsymbol{\theta}} = \mathbf{F}'\mathbf{y} - \boldsymbol{\Gamma}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \tag{35}$$

Note that, applying Lemma 1 in the sequel, it is easy to see that $\mathrm{var}(\tilde{\boldsymbol{\beta}}) - \mathrm{var}(\hat{\boldsymbol{\beta}}) \geq 0$. Thus, the covariance matrix of the BPE is always larger than or equal to that of the MLE (in the matrix ordering; see below). In particular, the BPE cannot be more efficient than the MLE when the underlying model is correct. However, this needs not be the case when the underlying model is misspecified. To compare the exact MSPE of the OBP and BLUP, let us consider a class of empirical best predictors (EBPs) that can be expressed as

$$\breve{\boldsymbol{\theta}} = \mathbf{F}'\mathbf{y} - \boldsymbol{\Gamma}(\mathbf{y} - \mathbf{X}\breve{\boldsymbol{\beta}}), \tag{36}$$

where $\breve{\boldsymbol{\beta}}$ is a weighted least squares (WLS) estimator of $\boldsymbol{\beta}$ expressed as

$$\breve{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}, \tag{37}$$

and $\mathbf{W}$ is a positive definite weighting matrix. Equation (36) is called EBP because it is obtained by replacing $\boldsymbol{\beta}$ in the BP (30) by a WLS estimator $\breve{\boldsymbol{\beta}}$. Note that the BPE and MLE are special cases of the WLS, hence the OBP and BLUP are special cases of the EBP. It follows that $\breve{\boldsymbol{\theta}} = [\mathbf{F}' - \boldsymbol{\Gamma}\{\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\}]\mathbf{y} \equiv \mathbf{L}\mathbf{y}$. By (28) and (29), it can be shown that

$$\begin{aligned}\mathrm{MSPE}(\breve{\boldsymbol{\theta}}) = {} & \mathrm{tr}\{\mathbf{R}'(\mathbf{G} - \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{G})\mathbf{R}\} \\ & + \boldsymbol{\mu}'\{\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\}' \\ & \times \boldsymbol{\Gamma}'\boldsymbol{\Gamma}\{\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\}\boldsymbol{\mu} \\ & + \mathrm{tr}\{(\mathbf{L} - \mathbf{B})\mathbf{V}(\mathbf{L} - \mathbf{B})'\}.\end{aligned} \tag{38}$$

The first term on the right-hand side of (38) does not depend on the EBP, that is, $\mathbf{W}$. As for the second term, we first use a matrix identity (e.g., Searle, Casella, and McCulloch 1992, p. 451): For any full rank matrix $\mathbf{K}$ (of compatible dimension) such that $\mathbf{K}'\mathbf{X} = 0$, we have

$$\mathbf{W} - \mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W} = \mathbf{K}(\mathbf{K}'\mathbf{W}^{-1}\mathbf{K})^{-1}\mathbf{K}'. \tag{39}$$

It follows that

$$\begin{aligned}&\{\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\}'\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\{\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\} \\ &\quad = \{\mathbf{W} - \mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\} \\ &\qquad \times \mathbf{W}^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\mathbf{W}^{-1}\{\mathbf{W} - \mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\} \\ &\quad = \mathbf{K}(\mathbf{K}'\mathbf{W}^{-1}\mathbf{K})^{-1}\mathbf{K}'\mathbf{W}^{-1} \\ &\qquad \times \boldsymbol{\Gamma}'\boldsymbol{\Gamma}\mathbf{W}^{-1}\mathbf{K}(\mathbf{K}'\mathbf{W}^{-1}\mathbf{K})^{-1}\mathbf{K}'.\end{aligned} \tag{40}$$

We then use the following lemma which is well known in the context of WLS (e.g., lemma 5.1 of Jiang 2010). A proof is given in the Appendix. Here, for symmetric matrices, $\mathbf{M} > 0$ (≥ 0) iff $\mathbf{M}$ is positive (nonnegative) definite, and $\mathbf{M}_1 \geq \mathbf{M}_2$ means that $\mathbf{M}_1 - \mathbf{M}_2 \geq 0$.

Lemma 1. For any symmetric matrices $\mathbf{V}_1, \mathbf{V}_2 > 0$ and full rank matrix $\mathbf{K}$, we have

$$(\mathbf{K}'\mathbf{V}_1\mathbf{K})^{-1}\mathbf{K}'\mathbf{V}_1\mathbf{V}_2\mathbf{V}_1\mathbf{K}(\mathbf{K}'\mathbf{V}_1\mathbf{K})^{-1} \geq (\mathbf{K}'\mathbf{V}_2^{-1}\mathbf{K})^{-1}. \tag{41}$$

By (41), with $\mathbf{V}_1 = \mathbf{W}^{-1}$ and $\mathbf{V}_2 = \boldsymbol{\Gamma}'\boldsymbol{\Gamma}$, we have

$$\begin{aligned}&\text{the right-hand side of (40)} \\ &\quad \geq \mathbf{K}\{\mathbf{K}'(\boldsymbol{\Gamma}'\boldsymbol{\Gamma})^{-1}\mathbf{K}\}^{-1}\mathbf{K}' \\ &\quad = \boldsymbol{\Gamma}'\boldsymbol{\Gamma} - \boldsymbol{\Gamma}'\boldsymbol{\Gamma}\mathbf{X}(\mathbf{X}'\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Gamma}'\boldsymbol{\Gamma},\end{aligned} \tag{42}$$

again using (39) (with $\mathbf{W}$ replaced by $\mathbf{\Gamma}'\mathbf{\Gamma}$) for the last identity. On the other hand, the lower bound on the right-hand side of (42) is attained by the OBP, which has $\mathbf{W} = \mathbf{\Gamma}'\mathbf{\Gamma}$.

Now consider the third term on the right-hand side of (38). Again by Lemma 1, we have

$$(\mathbf{L}-\mathbf{B})\mathbf{V}(\mathbf{L}-\mathbf{B})' = \mathbf{\Gamma X}(\mathbf{X}'\mathbf{W X})^{-1}\mathbf{X}'\mathbf{W V W X}(\mathbf{X}'\mathbf{W X})^{-1}\mathbf{X}'\mathbf{\Gamma}' \geq \mathbf{\Gamma X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Gamma}'. \tag{43}$$

On the other hand, the lower bound on the right-hand side of (43) is attained by the BLUP, which has $\mathbf{W} = \mathbf{V}^{-1}$. We summarize the results as a theorem.

Theorem 1. The MSPE of the EBP (36)–(37) can be expressed as

$$\mathrm{MSPE}(\check{\boldsymbol{\theta}}) = a_0 + \boldsymbol{\mu}'A_1(\mathbf{W})\boldsymbol{\mu} + \mathrm{tr}\{A_2(\mathbf{W})\}, \tag{44}$$

where $a_0 \geq 0$ is a constant that does not depend on the EBP; $A_1(\mathbf{W}), A_2(\mathbf{W}) \geq 0$ and depend on $\mathbf{W}$, hence the EBP. $\boldsymbol{\mu}'A_1(\mathbf{W})\boldsymbol{\mu}$ is minimized by the OBP, which has $\mathbf{W} = \mathbf{\Gamma}'\mathbf{\Gamma}$; $\mathrm{tr}\{A_2(\mathbf{W})\}$ is minimized by the BLUP, which has $\mathbf{W} = \mathbf{V}^{-1}$.

It is easy to verify that the example in Section 2 is a special case of Theorem 1.

Some important implications of Theorem 1 are the following. First note that the second term disappears when the mean of $\mathbf{y}$ is correctly specified, that is, $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ for some $\boldsymbol{\beta}$. Now suppose that the mean of $\mathbf{y}$ is misspecified. Then, it can be shown that $\boldsymbol{\mu}'A_1(\mathbf{W})\boldsymbol{\mu} \geq \lambda_{\min}(\mathbf{\Gamma}'\mathbf{\Gamma})|P_{\mathbf{X}^\perp}\boldsymbol{\mu}|^2$, where and hereafter $\lambda_{\min}$ denotes the smallest eigenvalue. Thus, provided that $\lambda_{\min}(\mathbf{\Gamma}'\mathbf{\Gamma})$ has a positive lower bound, the second term has the order $O(|P_{\mathbf{X}^\perp}\boldsymbol{\mu}|^2)$. On the other hand, the third term is typically $O(1)$. Thus, comparing the MSPE of OBP and BLUP, typically, the difference in their second terms is $O(|P_{\mathbf{X}^\perp}\boldsymbol{\mu}|^2)$ in favor of the OBP, while the difference in their third terms is $O(1)$ in favor of the BLUP. Overall, if $|P_{\mathbf{X}^\perp}\boldsymbol{\mu}|^2 \to \infty$, the OBP is expected to outperform the BLUP significantly in terms of the MSPE. On the other hand, even if the mean of $\mathbf{y}$ is correctly specified, in which case the second term vanishes, as long as the first term goes to infinity as the sample size increases [note that every term on the right-hand side of (44) depends on the sample size, n, so, for example, $a_0 = a_{0,n}$—we suppress the subscript n for notation simplicity], which is typically the case, one still has that the ratio of the MSPEs of the OBP and BLUP goes to one as the sample size increases; hence, the OBP and BLUP are expected to have similar performance asymptotically in terms of the MSPE.

5.2 Asymptotic Behavior of BPE

There have been studies examining the asymptotic behaviors of estimators under model misspecifications. For example, in the context of maximum likelihood estimation with iid observations, White (1982) showed that, when the underlying model is misspecified (therefore the true parameter vector does not exist), the MLE under the assumed model still converges almost surely to "something," and the "something" is the (unique) maximizer of the expected "log-likelihood" (considered as a function of the parameters), under some regularity conditions. Here "log-likelihood" refers to the logarithm of the misspecified likelihood function, which White called quasi-likelihood function. A similar approach can be used to study the asymptotic behavior of the BPE under a misspecified model.

We consider a general setting which includes the Fay–Herriot model, discussed, in particular, in Sections 1–3, and the nested-error regression model, discussed in Section 3, as special cases. Suppose that the BP of $\boldsymbol{\theta}$, a vector of mixed effects, under the assumed model, $\boldsymbol{\theta}_{\mathrm{BP}}$, depends on $\boldsymbol{\psi}$, a vector of parameters that may include $\boldsymbol{\beta}$ as well as some variance components. Also, suppose that $\mathrm{MSPE}(\boldsymbol{\theta}_{\mathrm{BP}})$ can be expressed as $\mathrm{E}\{Q(\mathbf{y}, \boldsymbol{\psi})\} \equiv M(\boldsymbol{\psi})$, where E denotes expectation under the true distribution of $\mathbf{y}$, either with respect to a true model (e.g., the case of Fay–Herriot model), or with respect to the sampling design (e.g., the case of nested-error regression model). Throughout this subsection, all the E, and later P, are understood in the same sense. We call $M(\boldsymbol{\psi})$ the *MSPE function*. The BPE of $\boldsymbol{\psi}$, denoted by $\tilde{\boldsymbol{\psi}}$, is the minimizer of $Q(\mathbf{y}, \boldsymbol{\psi})$ over $\boldsymbol{\psi} \in \Psi$, the parameter space for $\boldsymbol{\psi}$. The OBP of $\boldsymbol{\theta}$, $\tilde{\boldsymbol{\theta}}$, is $\boldsymbol{\theta}_{\mathrm{BP}}$ with $\boldsymbol{\psi}$ replaced by $\tilde{\boldsymbol{\psi}}$. Furthermore, we assume that

$$Q(\mathbf{y}, \boldsymbol{\psi}) = \sum_{i=1}^{m} Q_i(\mathbf{y}_i, \boldsymbol{\psi}), \tag{45}$$

where m is the number of small areas; $\mathbf{y}_i$ is the subvector of $\mathbf{y}$ corresponding to data collected from the ith small area, such that $\mathbf{y}_1, \ldots, \mathbf{y}_m$ are independent; and $Q_i(\mathbf{y}_i, \boldsymbol{\psi})$ is three-times continuously differentiable with respect to $\boldsymbol{\psi}$. In addition, the following regularity conditions are assumed.

A1. There exists a unique $\boldsymbol{\psi}_* \in \Psi^o$, the interior of Ψ, such that $M(\boldsymbol{\psi}_*) = \inf_{\boldsymbol{\psi}\in\Psi} M(\boldsymbol{\psi})$. Note that $\boldsymbol{\psi}_*$ may depend on the (joint) distribution of $\mathbf{y}_1, \ldots, \mathbf{y}_m$ as well as other quantities that may be involved in the definition of $M(\boldsymbol{\psi})$ (such as the $\mathbf{x}_i$'s in the Fay–Herriot model).

A2. One can differentiate $\mathrm{E}\{Q(\mathbf{y}, \boldsymbol{\psi})\}$ with respect to $\boldsymbol{\psi}$ under the expectation, that is, $(\partial/\partial\boldsymbol{\psi})\mathrm{E}\{Q(\mathbf{y}, \boldsymbol{\psi})\} = \mathrm{E}\{(\partial/\partial\boldsymbol{\psi}) Q(\mathbf{y}, \boldsymbol{\psi})\}$.

A3. As $m \to \infty$, we have:

(i) $\liminf \lambda_{\min}[m^{-1}\sum_{i=1}^m (\partial^2/\partial\boldsymbol{\psi}\,\partial\boldsymbol{\psi}')\mathrm{E}\{Q_i(\mathbf{y}_i, \boldsymbol{\psi}_*)\}] > 0$;

(ii) $\limsup m^{-1}\sum_{i=1}^m \mathrm{E}\{(\partial/\partial\psi_r)Q_i(\mathbf{y}_i, \boldsymbol{\psi}_*)\}^2 < \infty$;

(iii) $m^{-2}\sum_{i=1}^m \mathrm{E}\{(\partial^2/\partial\psi_r\,\partial\psi_s)Q_i(\mathbf{y}_i, \boldsymbol{\psi}_*)\}^2 \to 0$; and

(iv) $m^{-3/2}\sum_{i=1}^m \mathrm{E}\{\sup_{\boldsymbol{\psi}\in\bar{S}_\rho(\boldsymbol{\psi}_*)} |\partial^3 Q_i/\partial\psi_r\,\partial\psi_s\partial\psi_t|\} \to 0$ for some $\rho > 0$ and any $1 \leq r, s, t \leq q = \dim(\boldsymbol{\psi})$, where $\bar{S}_\rho(\boldsymbol{\psi}_*) = \{\boldsymbol{\psi} \in \Psi : |\boldsymbol{\psi} - \boldsymbol{\psi}_*| \leq \rho\}$.

Condition A1 requires identifiability of the parameter vector, $\boldsymbol{\psi}_*$, which assumes the role of the "true parameter vector." Condition A2 is a regularity condition that is often required in asymptotic theory (e.g., Lehmann and Casella 1998, p. 441). Condition (i) of A3 corresponds to an "information assumption," which is an extension of that (in the iid case) the sample size goes to infinity. The rest of the conditions in A3 are essentially moment conditions. These conditions are fairly mild in that they are satisfied in typical situations. We use an example to illustrate how to verify these conditions. Consider the Fay–Herriot model with A known, for simplicity. In this case, the minimizer of $M(\boldsymbol{\psi})$ has an explicit expression, $\boldsymbol{\psi}_* = (\mathbf{X}'\mathbf{\Gamma}^2\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Gamma}^2\mathrm{E}(\mathbf{y})$ [see (11)]. Clearly, $\boldsymbol{\psi}_*$ lies in the interior of the parameter space, R^p, where p is the dimension of $\boldsymbol{\beta}$. Thus, condition A1 is satisfied. Note that, in this case, $Q(\mathbf{y}, \boldsymbol{\psi})$ is given by the expression inside the expectation in (7). Thus, it is easy to verify that condition A2 is satisfied. Similarly, $Q_i(y_i, \boldsymbol{\psi}) = \mathrm{E}(B_i y_i - \theta_i)^2 + (1 - B_i)^2(\mathbf{x}_i'\boldsymbol{\beta})^2 - 2(1 - B_i)^2\mathbf{x}_i'\boldsymbol{\beta} y_i$. It follows that $m^{-1}\sum_{i=1}^m (\partial^2/\partial\boldsymbol{\psi}\,\partial\boldsymbol{\psi}')\mathrm{E}\{Q_i(y_i, \boldsymbol{\psi}_*)\} = 2m^{-1}\sum_{i=1}^m (1 - B_i)^2\mathbf{x}_i\mathbf{x}_i'$. Suppose that the D_i's are bounded away from zero. Then, it is easy to show that condition (i) of A3 is satisfied if $\liminf \lambda_{\min}(m^{-1}X'X) > 0$. The latter is a regularity condition often required in studying large sample properties of the least squares estimator (e.g., section 6.7 of Jiang 2010 and the references therein). For example, if $\mathbf{x}_i = (1, x_i)'$, where x_i is a scalar, then the latter condition is equivalent to $\liminf m^{-1}\sum_{i=1}^m (x_i - \bar{x})^2 > 0$, where $\bar{x} = m^{-1}\sum_{i=1}^m x_i$. The rest of the conditions in A3 are easy to verify; in particular, we have $\partial^3 Q_i/\partial\beta_j\,\partial\beta_k\,\partial\beta_l = 0$.

The following theorem states that, under a possibly misspecified model, the BPE is "$\sqrt{m}$-consistent" with respect to $\boldsymbol{\psi}_*$, the minimizer of the MSPE function.

Theorem 2. Under the assumptions A1–A3, there exists with probability tending to one as $m \to \infty$ a local minimizer, $\tilde{\boldsymbol{\psi}}$, of $Q(\mathbf{y}, \boldsymbol{\psi})$ in a neighborhood of $\boldsymbol{\psi}_*$, such that $\sqrt{m}(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_*) = O_{\mathrm{P}}(1)$. Thus, in particular, we have $\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_* \xrightarrow{\mathrm{P}} 0$ as $m \to \infty$.

Note that we do not write the last statement of Theorem 2 as $\tilde{\boldsymbol{\psi}} \xrightarrow{\mathrm{P}} \boldsymbol{\psi}_*$, because $\boldsymbol{\psi}_*$, although is nonrandom, may depend on m (see assumption A1). The proof is given in the Appendix. Theorem 2 plays an important role in the estimation of the area-specific MSPE of the OBP in the next section.

6. ESTIMATION OF AREA–SPECIFIC MSPE

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 and 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 and Rao 1995; Datta and Lahiri 2000; Jiang and Lahiri 2001; Datta, Rao, and 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. To be more specific, we focus on the Fay–Herriot model throughout this section. In this case, the only thing we know under our weak model assumption (2) is that $\tilde{\boldsymbol{\psi}}$, the BPE, is $\sqrt{m}$-consistent to $\boldsymbol{\psi}_*$ that is defined in Section 5.2, but $\boldsymbol{\psi}_*$ is not necessarily the true parameter vector. Even in the much simpler case in which $\boldsymbol{\psi}_*$ is known, hence one can replace $\tilde{\boldsymbol{\psi}}$ by $\boldsymbol{\psi}_*$ in the OBP, it is easy to see that the MSPE of $\tilde{\theta}_i$ is not a known function of $\boldsymbol{\psi}_*$ but depends on the unknown μ_i and $\mathrm{E}(y_i^2)$. It follows that, under the weak model assumption, the MSPE cannot be consistently estimated.

On the other hand, it is still possible to obtain a nearly unbiased estimator of the MSPE of $\tilde{\theta}_i$. First note that, instead of making the Taylor expansion at the point of the true parameter vector, as in the Prasad–Rao method, the expansion can be made at $\boldsymbol{\psi}_*$. However, the current weak assumption about the mean functions makes it difficult to carry out the standard arguments of Prasad and Rao, which make use of such assumptions as $\mathrm{E}(y_i) = \mathbf{x}_i'\boldsymbol{\beta}$ and $\mathrm{E}\{(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2\} = A + D_i$, $1 \le i \le m$, where $\boldsymbol{\psi} = (\boldsymbol{\beta}', A)'$ is the true parameter vector. Nevertheless, the asymptotic property of the BPE (Theorem 2), and a different technique allow us to obtain a second-order unbiased MSPE estimator. First introduce some notation. Let $\tilde{B}_i$ be B_i with A replaced by $\tilde{A}$. Define $\tilde{\mathbf{M}}_i = \mathrm{diag}\{x_i, (\tilde{A} + D_i)^{-1}\}$, $\tilde{\mathbf{U}}_i = (u_i, u_{m+i})'$, where $u_i = y_i - \mathbf{x}_i'\tilde{\boldsymbol{\beta}}$ and $u_{m+i} = (y_i - \mathbf{x}_i'\tilde{\boldsymbol{\beta}})^2 - (\tilde{A} + D_i)$. Let $\tilde{\mathbf{W}}_i = (\tilde{\mathbf{W}}_{i,ab})_{a,b=1,2}$ with $\tilde{\mathbf{W}}_{i,11} = \mathbf{x}_i\mathbf{x}_i'$, $\tilde{\mathbf{W}}_{i,12} = 2(\tilde{A} + D_i)^{-1}(y_i - \mathbf{x}_i'\tilde{\boldsymbol{\beta}})\mathbf{x}_i$, $\tilde{\mathbf{W}}_{i,21} = \tilde{\mathbf{W}}_{i,12}'$, and $\tilde{\mathbf{W}}_{i,22} = (\tilde{A} + D_i)^{-2}\{3(y_i - \mathbf{x}_i'\tilde{\boldsymbol{\beta}})^2 - (\tilde{A} + D_i)\}$. Let $\tilde{\mathbf{f}}_i = -2(1 - \tilde{B}_i)^2\tilde{\mathbf{M}}_i\tilde{\mathbf{U}}_i$ and $\tilde{\mathbf{G}}_2 = 2\sum_{j=1}^m (1 - \tilde{B}_j)^2(\tilde{\mathbf{W}}_j - \tilde{\boldsymbol{\Delta}}_j)$, where $\tilde{\boldsymbol{\Delta}}_j = \mathrm{diag}\{0, \ldots, 0, (\tilde{A} + D_j)^{-1}\}$ (p zeros; p is the dimension of $\boldsymbol{\beta}$), and let $\tilde{\mathbf{h}}_2'$ be the last row of $\tilde{\mathbf{G}}_2^{-1}$. It is shown in the Appendix that the following is a second-order unbiased estimator of the MSPE of $\tilde{\theta}_i$, the OBP of θ_i,

$$\widetilde{\mathrm{MSPE}}(\tilde{\theta}_i) = (\tilde{\theta}_i - y_i)^2 + D_i(2\tilde{B}_i - 1) + 2(1 - \tilde{B}_i)^2\tilde{\mathbf{h}}_2'\tilde{\mathbf{f}}_i + 4D_i(1 - \tilde{B}_i)^3\,\mathrm{tr}(\tilde{\mathbf{G}}_2^{-1}\tilde{\mathbf{W}}_i). \quad (46)$$

Some simulation studies are carried out to evaluate the finite-sample performance of the area-specific MSPE estimators. The estimator (46) is compared with a naive estimator,

$$\widetilde{\mathrm{MSPE}}_{\mathrm{n}}(\tilde{\theta}_i) = (\tilde{\theta}_i - y_i)^2 + D_i(2\tilde{B}_i - 1) \quad (47)$$

(see the Appendix for the derivation). Note that $\widetilde{\mathrm{MSPE}}_{\mathrm{n}}(\tilde{\theta}_i)$ is the sum of the first two terms in the expression of $\widetilde{\mathrm{MSPE}}(\tilde{\theta}_i)$. While $\widetilde{\mathrm{MSPE}}(\tilde{\theta}_i)$ has a bias of the order $o(m^{-1})$, $\widetilde{\mathrm{MSPE}}_{\mathrm{n}}(\tilde{\theta}_i)$ has a bias of the order $O(m^{-1})$. Consider the simulated example of Section 4.2 with $m = 23$. In addition to MSPE estimation of the OBP, we consider, as a comparison, estimation of the (area-specific) MSPE of EBLUP-4, which appears to perform the best among the EBLUPs in this case, according to our simulation results in Section 4.2. A well-known second-order unbiased MSPE estimator for EBLUP-4 is the Prasad–Rao (P–R) estimator, given by $\widehat{\mathrm{MSPE}}(\hat{\theta}_i) = g_{1,i}(\hat{A}) + g_{2,i}(\hat{A}) + 2g_{3,i}(\hat{A})$, where $\hat{\theta}_i$ is the EBLUP-4 for θ_i, $\hat{A}$ is the P–R estimator of A, and the expressions of $g_{j,i}(A), j = 1, 2, 3$, are given by equations (5.16)–(5.19) of Prasad and Rao (1990). The P–R MSPE estimator is compared with a naive MSPE estimator, $\widehat{\mathrm{MSPE}}_{\mathrm{n}}(\hat{\theta}_i) = g_{1,i}(\hat{A}) + g_{2,i}(\hat{A})$. A standard measure of the bias of an MSPE estimator is the percentage relative bias (%RB), defined by %RB = 100 × {(Mean of MSPE Estimator − True MSPE)/True MSPE} for each small area, where Mean of MSPE Estimator is obtained by the average of the MSPE estimator over $K = 5000$ simulations, and the True MSPE is approximated by the average of $(\check{\theta}_i - \theta_i)^2$ over the same 5000 simulations, where $\check{\theta}_i$ is either the OBP or the EBLUP-4 of θ_i. The %RBs are reported in Table 3. The corresponding Monte Carlo percentage relative SEs (%RSEs) are reported in Table 3 of the Appendix. It appears

Table 3. %RB of second-order unbiased and naive MSPE estimators

	Case I, OBP		Case I, EBLUP-4		Case II, OBP		Case II, EBLUP-4	
Area	$\widetilde{\mathrm{MSPE}}$	$\widetilde{\mathrm{MSPE}}_{\mathrm{n}}$	$\widehat{\mathrm{MSPE}}$	$\widehat{\mathrm{MSPE}}_{\mathrm{n}}$	$\widetilde{\mathrm{MSPE}}$	$\widetilde{\mathrm{MSPE}}_{\mathrm{n}}$	$\widehat{\mathrm{MSPE}}$	$\widehat{\mathrm{MSPE}}_{\mathrm{n}}$
1	−1.9	−36.3	−56.5	−64.1	2.4	−81.5	−24.6	−38.6
2	−8.0	−39.4	−55.4	−63.4	−7.9	−79.5	−25.0	−39.3
3	−4.3	−47.4	−44.0	−54.2	−16.7	−70.2	−25.3	−40.3
4	−6.6	−28.5	−50.8	−59.9	−11.6	−78.0	−23.5	−38.9
5	2.6	−98.9	−44.7	−49.4	−0.1	−148.9	9.2	−0.1
6	−6.1	−23.4	−46.0	−57.3	−9.6	−56.5	−23.8	−41.5
7	−4.1	−21.7	−66.8	−73.6	−3.1	−63.3	−20.4	−37.3
8	−14.1	−43.5	−22.6	−38.6	−7.8	−48.8	−22.7	−41.1
9	−1.3	−12.5	−81.8	−85.6	8.6	−71.2	−11.4	−28.2
10	−4.8	−36.2	−33.9	−46.9	−5.5	−81.7	−23.2	−40.3
11	−9.9	−33.6	−19.2	−37.4	−0.5	−36.4	−23.0	−43.3
12	−8.8	−28.9	−12.9	−33.2	−5.3	−35.9	−25.5	−45.8
13	−5.8	−25.6	−38.9	−53.0	−3.0	−37.3	−20.0	−41.0
14	−2.6	−16.5	−49.3	−62.6	4.1	−28.0	−17.6	−42.4
15	−6.3	−19.6	−46.3	−60.3	−9.7	−31.8	−15.1	−40.4
16	−3.9	−18.2	−46.3	−59.4	−0.6	−52.6	−13.5	−37.4
17	0.8	−6.2	−64.4	−74.8	−2.0	−27.1	−9.4	−38.3
18	1.7	−5.2	−64.0	−74.8	4.1	−22.9	−2.9	−35.0
19	−4.9	−12.7	−36.7	−56.1	−2.6	−20.4	−9.0	−41.9
20	−1.4	−10.5	−35.1	−56.2	−5.9	−20.0	−3.1	−40.3
21	−9.0	−22.9	−2.7	−33.2	−3.0	−26.6	−8.1	−43.1
22	−5.3	−12.2	−18.9	−47.6	−9.3	−19.8	−2.5	−44.0
23	−2.0	−9.4	−33.6	−57.0	0.1	−23.1	23.0	−28.0

that the second-order unbiased OBP MSPE estimator has more "correction power," with respect to the corresponding naive estimator, than the P–R second-order unbiased MSPE estimator for EBLUP-4, especially under Case I. This makes sense because, as noted, the Prasad–Rao method of MSPE estimation is developed under the assumption that the underlying model is correct, while this is not the case in Case I. On the other hand, our second-order unbiased MSPE estimator (46) is derived without assuming that the underlying model is correct; in fact, it seems to perform equally well in both cases. However, the comparisons made here are not entirely fair, because the MSPE estimators for OBP and EBLUP-4 are, of course, estimators of different things, and the definitions of the so-called naive (MSPE) estimators are also somewhat subjective.

On the other hand, unlike the P–R MSPE estimator, the MSPE estimator (46) is not guaranteed to be nonnegative. Alternatively, we may use the following bootstrap method, whose derivation is given in the Appendix, to obtain an MSPE estimator that is guaranteed nonnegative. Let $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_i)_{1\le i\le m}$ denote the vector of OBP. First generate $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(L)}$ independently from the $\mathrm{N}(\tilde{\boldsymbol{\theta}}, \mathbf{D})$ distribution, where $\mathbf{D} = \mathrm{diag}(D_i, 1 \le i \le m)$. Let $\tilde{\theta}_i^{(l)}$ be the OBP for θ_i based on $\mathbf{y}^{(l)}$, $1 \le i \le m$, $1 \le l \le L$. Then, the bootstrap estimator of $\mathrm{MSPE}(\tilde{\theta}_i)$ is given by

$$\widetilde{\mathrm{MSPE}}_{\mathrm{b}}(\tilde{\theta}_i) = \frac{1}{L}\sum_{l=1}^{L}\{\tilde{\theta}_i^{(l)} - \tilde{\theta}_i\}^2. \tag{48}$$

Another simulation study is carried out to evaluate the performance of the MSPE estimator (48). The results are presented in the Appendix. It appears that the bootstrap MSPE estimator does not perform as well as the second-order unbiased MSPE estimator (46) in terms of the bias reduction. In fact, the bootstrap MSPE estimator is only first-order unbiased (see Section 8 for further discussion). On the other hand, the bootstrap MSPE estimator seems to be more robust than the naive MSPE estimator (47), especially in Case II, in terms of poorly estimated MSPEs. Once again, the main advantage of the MSPE estimator (48) over the MSPE estimators (46) and (47) is that it is guaranteed nonnegative (note that the naive MSPE estimator is more likely to take negative values). While theoretical properties of the bootstrap MSPE estimator will be studied in subsequent research, our general recommendation is to use the MSPE estimator (46), if it is nonnegative; otherwise, use the MSPE estimator (48). Similar approaches have been used, for example, by Hall and Maiti (2006), to deal with the situation where their nonparametric bootstrap MSPE estimator is negative.

7. HOSPITAL DATA REVISITED

Following Section 4.2, we consider OBP and EBLUP-4 under the Q–O model. The latter appears to perform the best among the EBLUPs in this case based on the simulation results. Furthermore, we obtain the MSPE estimates (46) for the OBP and the P–R MSPE estimates for the EBLUP-4. For six out of the 23 areas (area #3, 6, 7, 11, 20, and 23) the estimates (46) are negative. Following the recommendation of Section 6 (near the end), for those areas we use the bootstrap MSPE estimates (48) with $L = 100$ for the OBP. The BPE for the parameters are, approximately, $\tilde{\beta}_0 = -0.084$, $\tilde{\beta}_1 = 4.614$, $\tilde{\beta}_2 = -16.045$, $\tilde{d} = 0.698$, and $\tilde{A} = 3.4 \times 10^{-4}$; the corresponding estimates for EBLUP-4 are, approximately, $\hat{\beta}_0 = -0.040$, $\hat{\beta}_1 = 3.723$, $\hat{\beta}_2 = -12.745$, $\hat{d} = 0.580$, and $\hat{A} = 3.6 \times 10^{-4}$. The OBP and EBLUP-4 are reported in Table 4 with the square

Table 4. The hospital data, EBLUP-4, OBP with measure of uncertainty

Area	y_i	x_i	$\sqrt{D_i}$	OBP ($\widetilde{\text{RMSPE}}$)	EBLUP-4 ($\widehat{\text{RMSPE}}$)
1	0.302	0.112	0.055	0.239 (0.060)	0.226 (0.026)
2	0.140	0.206	0.053	0.181 (0.019)	0.181 (0.027)
3	0.203	0.104	0.052	0.220 (0.017)	0.209 (0.026)
4	0.333	0.168	0.052	0.249 (0.085)	0.238 (0.026)
5	0.347	0.337	0.047	0.347 (0.047)	0.347 (0.049)
6	0.216	0.169	0.046	0.234 (0.016)	0.224 (0.026)
7	0.156	0.211	0.046	0.172 (0.020)	0.175 (0.028)
8	0.143	0.195	0.046	0.197 (0.045)	0.193 (0.026)
9	0.220	0.221	0.044	0.162 (0.058)	0.170 (0.031)
10	0.205	0.077	0.044	0.180 (0.017)	0.176 (0.028)
11	0.209	0.195	0.042	0.206 (0.015)	0.203 (0.026)
12	0.266	0.185	0.041	0.228 (0.031)	0.222 (0.026)
13	0.240	0.202	0.041	0.201 (0.031)	0.200 (0.026)
14	0.262	0.108	0.036	0.234 (0.024)	0.224 (0.027)
15	0.144	0.204	0.036	0.180 (0.032)	0.179 (0.027)
16	0.116	0.072	0.035	0.154 (0.040)	0.152 (0.029)
17	0.201	0.142	0.033	0.236 (0.032)	0.224 (0.027)
18	0.212	0.136	0.032	0.238 (0.020)	0.226 (0.027)
19	0.189	0.172	0.031	0.223 (0.031)	0.214 (0.026)
20	0.212	0.202	0.029	0.199 (0.014)	0.198 (0.026)
21	0.166	0.087	0.029	0.187 (0.017)	0.181 (0.026)
22	0.173	0.177	0.027	0.212 (0.039)	0.204 (0.025)
23	0.165	0.072	0.025	0.165 (0.017)	0.163 (0.027)

roots of the corresponding MSPE estimates (RMSPE; in the parentheses) used as measures of uncertainty.

If the assumed Q–O model is correct, then the simulation results of Section 4.2 suggest that EBLUP-4 may work better in this case; on the other hand, if the Q–O model is misspecified, the OBP may have an edge. While we may never know for sure whether the Q–O model is correct, it is very possible, as is usually the case, that the underlying mean function is (still) misspecified, to some extend. For example, a plot (see Figure 2 in the Appendix) suggests that the fit (by either method) is relatively poor for the middle part of the x range, where a quadratic mean function is fitted.

Also note that the P–R MSPE estimates are almost constant across the areas, with the exception of area #5 that corresponds to the outlying case. As for the OBP, areas with the largest estimated MSPEs correspond to those for which the quadratic curve fits the data relatively poorly, such as areas #1 and #4, but not area #5, for which the Q–O model fits the data well (see Figure 2 in the Appendix). This makes sense because the OBP MSPE estimates take into account the potential bias caused by model misspecification—and this is also with respect to a specific small area.

8. 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. For example, the Fay–Herriot model under the very weak assumption (2) and the design-based considerations in Section 3. One may also consider nonparametric models (e.g., Opsomer et al. 2008) that are less restrictive than parametric ones. Thus, there is some flexibility in choosing the broader model. 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. Our approach is to develop a prediction 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 definition of the MSPE is based on the squared Euclidean distance between the vector of small area means, $\boldsymbol{\theta} = (\theta_i)_{1\le i\le m}$, and its vector-valued predictor, $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_i)_{1\le i\le m}$, that is, $|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}|^2 = \sum_{i=1}^m (\tilde{\theta}_i - \theta_i)^2$. From another point of view, the MSPE may be regarded as an "overall" MSPE for all the small areas, in which equal weights are given to all the small areas. It may be argued, however, that this is not necessarily the "best" way to assign the weights. For example, it may be reasonable to assign larger weights to areas where the variances of the direct estimators are large. Interestingly, the BPE (9) does exactly this (unlike the MLE). One may also argue that larger weights should be assigned to more important areas. This certainly brings up an interesting problem of future research.

In this article, the focus is misspecification of the mean function. There are, of course, other types of possible model misspecifications. For example, the underlying assumption for the random effects in the Fay–Herriot model is that they are independent and normally distributed. Both the normality and the independence assumptions could be violated, in practice. In particular, there may be spatial correlations among the random effects for the neighboring areas. Consider the general linear mixed model (27), where $\mathbf{v} \sim \mathrm{N}(0, \mathbf{G})$, $\mathbf{e} \sim \mathrm{N}(0, \boldsymbol{\Sigma})$, and $\mathbf{v}, \mathbf{e}$ are independent. It can be shown that the normality assumption is not used in the derivation of the OBP in Section 5.1, neither is the assumption that $\mathbf{G}$ is known, provided that $\boldsymbol{\Sigma}$ is known. In fact, the BPE for $\boldsymbol{\beta}$ and $\mathbf{G}$ are obtained by minimizing $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Gamma}'\boldsymbol{\Gamma}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - 2\,\mathrm{tr}(\boldsymbol{\Gamma}'\boldsymbol{\Sigma})$ (note that $\mathbf{G}$ is involved in $\boldsymbol{\Gamma}$). Thus, without assuming that $\mathbf{G}$ is known, which could be misspecified due to spatial correlations, for example, one could estimate $\mathbf{G}$ using the BPE, which is derived without using the assumption about $\mathbf{G}$ (and normality). However, at this point, we do not know anything about the performance of the OBP when $\mathbf{G}$ is misspeficied, or when the random effects and errors are not normal. Also, we need to deal with the more practical situation that $\boldsymbol{\Sigma}$ is unknown.

The parametric bootstrap MSPE estimator proposed at the end of Section 6 is not expected to be second-order unbiased. To make it second-order unbiased one needs to do, perhaps, a double-bootstrap bias correction, as in Hall and Maiti (2006). The procedure is computationally intensive, and the bias-corrected MSPE estimator may not be guaranteed nonnegative. It is therefore an interesting problem to find a second-order unbiased MSPE estimator that is always nonnegative. The Prasad–Rao MSPE estimator for the EBLUP, for example, has this nice property.

In this article, we show that the OBP is more robust than the EBLUP with respect to misspecification of the mean function, considered as part of the underlying model. However, this is, by no means, to undermine the importance of statistical modeling. A good model is the key in the business of "borrowing strength" in SAE. In fact, in case of a serious model misspecification, the assumed model may not be useful in the first place. For example, consider the simulated example of Section 4.1. It can be shown that, in this case, the variance A is seriously overestimated by all the methods being considered (EBLUPs or OBP). On the other hand, the expression of the BP, that is, (5), shows that, if the estimate of A is significantly larger than the value of D_i, the BP will be close to the direct estimator. It follows that the assumed model is not making much of a difference as compared to the direct estimator (hence not much strength can be borrowed by using such a model). As indicated (see the beginning of Section 4.1), this example is used to demonstrate the theoretical properties of the OBP and its comparison with the EBLUPs. In practice, however, such a serious misspecification of the mean function would most likely be detected, and hence corrected, at least to some extent, if one is indeed careful about the modeling part of the SAE. The latter includes model testing, model diagnostics, and model selection (e.g., Jiang, Nguyen, and Rao 2010). On the other hand, in some cases, it is possible to "control" the potential damage of poor variance estimation. We show how this can be done in another simulated example (see Section 4.2), which imitates a more practical situation. The idea applies to situations where one or more nonbinary covariates are available, but not to the case of Section 4.1. Nevertheless, from a theoretical standpoint, it is interesting to know what would happen if no "damage control" is done regarding the variance estimation, and this is shown in Section 4.1.

SUPPLEMENTARY MATERIALS

Appendix to Best Predictive SAE: This supplementary appendix contains some technical results and further details of the article. The materials include A.1. *A plot mentioned in Section* 1, A.2. *Some derivation in Section* 2, A.3. *OBP under the nested-error regression model*, A.4. *Derivation of* (26), A.5. *Monte Carlo errors of simulations*, A.6. *Proof of Lemma* 1, A.7. *Proof of Theorem* 2, A.8. *Estimation of area-specific MSPE under the Fay–Herriot model*, A.9. *The bootstrap MSPE estimator*, and A.10. *A plot for Section* 7. (jnr.suppl.pdf)

[Received April 2010. Revised January 2011.]

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*, 80, 28–36. [734,735]

Das, K., Jiang, J., and Rao, J. N. K. (2004), "Mean Squared Error of Empirical Predictor," *The Annals of Statistics*, 32, 818–840. [733]

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. [742]

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. [735,742]

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 the American Statistical Association*, 74, 269–277. [732,735]

Ganesh, N. (2009), "Simultaneous Credible Intervals for Small Area Estimation Problems," *Journal of Multivariate Analysis*, 100, 1610–1621. [732]

Hall, P., and Maiti, T. (2006), "Nonparametric Estimation of Mean-Squared Prediction Error in Nested-Error Regression Models," *The Annals of Statistics*, 34, 1733–1750. [743,745]

Jiang, J. (2007), *Linear and Generalized Linear Mixed Models and Their Applications*, New York: Springer. [732]

——— (2010), *Large Sample Techniques for Statistics*, New York: Springer. [740,742,745]

Jiang, J., and Lahiri, P. (2001), "Empirical Best Prediction for Small Area Inference With Binary Data," *Annals of the Institute Statistical Mathematics*, 53, 217–243. [742]

——— (2006), "Mixed Model Prediction and Small Area Estimation" (with discussion), *TEST*, 15, 1–96. [732]

Jiang, J., Nguyen, T., and Rao, J. S. (2010), "Fence Method for Nonparametric Small Area Estimation," *Survey Methodology*, 36, 3–11. [737]

Lahiri, P., and Rao, J. N. K. (1995), "Robust Estimation of Mean Squared Error of Small Area Estimators," *Journal of the American Statistical Association*, 90, 758–766. [742]

Lehmann, E. L., and Casella, G. (1998), *Theory of Point Estimation* (2nd ed.), New York: Springer. [741]

Morris, C. N., and Christiansen, C. L. (1995), "Hierarchical Models for Ranking and for Identifying Extremes With Applications," in *Bayes Statistics 5*, Oxford: Oxford University Press. [732]

Opsomer, J. D., Breidt, F. J., Claeskens, G., Kauermann, G., and Ranalli, M. G. (2008), "Nonparametric Small Area Estimation Using Penalized Spline Regression," *Journal of the Royal Statistical Society, Ser. B*, 70, 265–286. [744]

Pfeffermann, D., and Nathan, G. (1981), "Regression Analysis of Data From a Cluster Sample," *Journal of the American Statistical Association*, 76, 681–689. [735]

Prasad, N. G. N., and Rao, J. N. K. (1990), "The Estimation of Mean Squared Errors of Small Area Estimators," *Journal of the American Statistical Association*, 85, 163–171. [733,735,742]

Rao, J. N. K. (2003), *Small Area Estimation*, New York: Wiley. [732,735]

Robinson, G. K. (1991), "That BLUP Is a Good Thing: The Estimation of Random Effects" (with discussion), *Statistica Sinica*, 6, 15–51. [740]

Searle, S. R., Casella, G., and McCulloch, C. E. (1992), *Variance Components*, New York: Wiley. [740]

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25. [741]