Statistica Sinica **32** (2022), 2217-2240
doi:https://doi.org/10.5705/ss.202020.0325

THAT PRASAD-RAO IS ROBUST: ESTIMATION OF MEAN SQUARED PREDICTION ERROR OF OBSERVED BEST PREDICTOR UNDER POTENTIAL MODEL MISSPECIFICATION

Xiaohui Liu[1], Haiqiang Ma[1] and Jiming Jiang[1,2]

[1]*Jiangxi University of Finance and Economics*
and [2]*University of California, Davis*

Abstract: This study examines a measure of uncertainty for robust small area estimation (SAE). We consider the estimation of the mean squared prediction error (MSPE) of the observed best predictor (OBP) in SAE under the Fay-Herriot model with potential model misspecification. Previously, it was thought that the traditional Prasad-Rao (PR) linearization method could not be used, because it is derived under the assumption that the underlying model is correctly specified. However, we show that when it comes to estimating the unconditional MSPE, the PR estimator, derived for estimating the MSPE of the OBP, assuming that the underlying model is correct, remains first-order unbiased, even when the underlying model is misspecified in its mean function. A second-order unbiased estimator of the MSPE is derived by modifying the PR MSPE estimator. The PR and modified PR estimators also have much smaller variation than that of existing MSPE estimators for the OBP. The theoretical findings are supported by empirical results, including simulation studies and real-data applications.

Key words and phrases: Fay-Herriot model, model misspecification, observed best prediction, robustness, second-order unbiasedness, small area estimation

1. Introduction

Robust small area estimation (SAE; e.g., Rao and Molina (2015)) has received considerable attention in recent studies; see, for example, Sinha and Rao (2009), Jiang, Nguyen and Rao (2011), and Jiongo, Haziza and Duchesne (2013). In particular, Jiang, Nguyen and Rao (2011) introduced the observed best prediction (OBP) method which is known to be more robust against model misspecification than is the traditional empirical best linear unbiased prediction (EBLUP) method. See Pfeffermann (2013), Jiang, Nguyen and Rao (2015), Chen, Jiang and Nguyen (2015), and Jiang and Rao (2020) for reviews and extensions. The

Corresponding author: Jiming Jiang, University of California, Davis, Davis, CA 95616, USA. E-mail: jimjiang@ucdavis.edu.

robustness of the OBP is achieved by entertaining two models: an assumed model (e.g., a linear mixed model), which is used to derive the best predictor (BP), and a broader model, which is used to derive estimators for the parameters involved in the BP. The assumed model is more useful, owing to its simplicity and the direct use of available covariates. However, owing to the specificity of the model, the assumed model is likely to be misspecified. The broader model, on the other hand, is almost always correct, but is not useful in terms of using the available covariates. The broader model is only used to derive estimators for the parameters in the BP in a way that is not affected by the model misspecification. Note that the difference between the OBP and the traditional EBLUP is in how the unknown parameters in the BP are estimated (see below).

Nevertheless, the weak assumption of the broader model makes it difficult to assess the uncertainty associated with the OBP. This is because the OBP is derived by taking into account the potential model misspecification. Therefore, to derive a measure of uncertainty, the potential model misspecification also needs to be considered. More importantly, it is desirable to evaluate the uncertainty due to the potential model misspecification. A standard measure of uncertainty is the mean squared prediction error (MSPE). In this study, we focus on the area-level model, or Fay-Herriot model (Fay and Herriot (1979)), arguably the most widely used model in SAE. The model can be expressed in terms of a linear mixed model (LMM) as

$$y_i = x_i'\beta + v_i + e_i, \quad i = 1, \ldots, m, \tag{1.1}$$

where x_i is a vector of known covariates, β is a vector of unknown regression coefficients (the fixed effects), v_i is an area-specific random effect, and e_i is a sampling error. It is assumed that v_i and e_i 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 variance, D_i, is known, for $1 \le i \le m$. The problem of interest is to estimate the small area means, which, under the assumed model, are $\theta_i = x_i'\beta + v_i$, for $1 \le i \le m$. If β and A are known, the BP for θ_i under the assumed model is

$$\tilde{\theta}_i = x_i'\beta + \frac{A}{A + D_i}(y_i - x_i'\beta) = B_i y_i + (1 - B_i)x_i'\beta, \tag{1.2}$$

where $B_i = A/(A + D_i)$, for $1 \le i \le m$. Note that the BP is in the sense of minimizing the MSPE under model (1.1), denoted by $\mathrm{MSPE}_1 = \mathrm{E}_1(\tilde{\theta}_i - \theta_i)^2$. Here the expectation E_1 is with respect to model (1.1), that is, assuming that (1.1) holds, and that β and A involved in $\tilde{\theta}_i$ are the true parameters under model

(1.1). This is different to the true MSPE and expectation, denoted by MSPE and E (without the subscript 1), respectively, introduced below under a broader model.

Let $\theta = (\theta_i)_{1\le i\le m}$ denote a vector of the true small area means, which may or may not be expressed as $x_i'\beta + v_i$. If in (1.2) β and A are treated as fixed and unknown parameters, Jiang, Nguyen and Rao (2011) showed that the MSPE of $\tilde\theta = (\tilde\theta_i)_{1\le i\le m}$ has the following expression:

$$\begin{aligned}\mathrm{MSPE}(\tilde\theta) &= \mathrm{E}(|\tilde\theta - \theta|^2)\\ &= \mathrm{E}\{(y - X\beta)'\Gamma^2(y - X\beta) + 2A\mathrm{tr}(\Gamma) - \mathrm{tr}(D)\},\end{aligned} \tag{1.3}$$

where $X = (x_i')_{1\le i\le m}$, $y = (y_i)_{1\le i\le m}$, $\Gamma = \mathrm{diag}(1 - B_i, 1 \le i \le m)$, and $D = \mathrm{diag}(D_i, 1 \le i \le m)$. In (1.3), the expectation is with respect to the true distribution of y under the following broader model:

$$y_i = \mu_i + v_i + e_i, \quad i = 1, \ldots, m, \tag{1.4}$$

where μ_i is completely unknown, and v_i and e_i are as in (1.1). Note that (1.4) is much broader than (1.1). It follows that the true small area means can be expressed as $\theta_i = \mu_i + v_i$, for $1 \le i \le m$. The idea is to find estimators of $\boldsymbol{\beta}$ and A that minimize the expression inside the expectation on the right side of (1.3), that is,

$$Q(\beta, A) = (y - X\beta)'\Gamma^2(y - X\beta) + 2A\mathrm{tr}(\Gamma) - \mathrm{tr}(D).$$

These are known as the best predictive estimators (BPE), and are denoted by $\hat\beta$ and $\hat A$, respectively. The BPE is different to, for example, the standard maximum likelihood estimator (MLE). We illustrate the difference using a simple example.

Example 1. For simplicity, assume that A is known. Then, under model (1.1), the BPE of β has the expression

$$\hat\beta = \left\{\sum_{i=1}^m \left(\frac{D_i}{A + D_i}\right)^2 x_i x_i'\right\}^{-1} \sum_{i=1}^m \left(\frac{D_i}{A + D_i}\right)^2 x_i y_i.$$

In comparison, the MLE of β has the expression

$$\tilde\beta = \left(\sum_{i=1}^m \frac{x_i x_i'}{A + D_i}\right)^{-1} \sum_{i=1}^m \frac{x_i y_i}{A + D_i}.$$

Comparing the two expressions, we see that both estimators are weighted averages of the data; the only difference is in how the weights are assigned. Whereas the

MLE gives more weight to areas with a smaller sampling variance, D_i, the BPE dose the opposite, giving more weight to areas with a larger sampling variance.

As noted, the only difference between the OBP and the EBLUP is in how the model parameters in the BP are estimated; in fact, both are special cases of the empirical best predictor (EBP), that is, the BP with the model parameters replaced by some estimators. The OBP of θ_i, denoted by $\hat{\theta}_i$, is obtained using (1.2), with β and A replaced by their BPEs, for $1 \leq i \leq m$. The OBP is more robust to a misspecification of the assumed model than is the EBLUP, because the estimators used in the OBP are derived under an objective function, that is, (1.3), where the E does not depend on the assumed model. In contrast, the estimators used in the EBLUP are derived using an objective function that depends on the assumed model, such as the log-likelihood or restricted log-likelihood (e.g., Jiang (2007), sec. 1.3). This difference is illustrated in Example 1. As explained in Jiang, Nguyen and Rao (2011), this is why the OBP, which corresponds to the BPE, is more robust to model misspecification than is the EBLUP, which in this case corresponds to the MLE, in terms of predictive performance.

The main focus of this study is a measure of uncertainty for the OBP, namely, its MSPE. Here, the MSPE refers to that under the true underlying model, which in our case is model (1.4). The estimation of area-specific MSPEs has been studied extensively in the SAE literature. See, for example, Datta et al. (2011), Pfeffermann (2013), and Rao and Molina (2015) for recent reviews. It is desirable to obtain a second-order unbiased estimator of the MSPE in the sense that the bias of the MSPE estimator is $o(m^{-1})$, where m is the number of small areas from which data are collected. This is challenging, both analytically and computationally; see, for example, Prasad and Rao (1990), Jiang, Lahiri and Wan (2002), Hall and Maiti (2006), Datta et al. (2011), Rao and Molina (2015), Jiang (2017), and Jiang and Torabi (2020). Note that these works assume that the assumed model is correct; that is, model (1.1) holds. The problem is even harder when the assumed model is potentially misspecified, which is the situation we are dealing with.

A well-known method for obtaining a second-order unbiased MSPE estimator is the Prasad-Rao (PR) linearization method (Prasad and Rao (1990)). The method is developed under the assumption that the underlying model is correct. In fact, the assumed model is heavily used in the derivation of the PR MSPE estimator. Given that, it is somewhat surprising that in the case of a model misspecification, the PR MSPE estimator for the OBP (which is the EBP, with the model parameters estimated by the BPE) is still mostly correct. In fact, the

PR MSPE estimator remains first-order unbiased in the sense that the bias of the estimator is $O(m^{-1})$, even if the underlying model is misspecified in its mean function. Note that the order of the area-specific MSPE is typically $O(1)$. It follows that the PR MSPE estimator is asymptotically unbiased. Furthermore, the PR MSPE estimator can be modified to achieve second-order unbiasedness, again under potential model misspecification in the mean function.

Note that the robust feature of the PR estimator has been previously found in a different aspect. Lahiri and Rao (1995) considered the Fay-Herriot model, (1.1), and showed that the PR estimator of the MSPE of the EBLUP, derived under the normality of both v_i and e_i, remains second-order unbiased when the normality assumption about v_i is relaxed to a certain moment condition. However, the normality of e_i is still needed. The phrase, "That Prasad-Rao is Robust", in the title of the current paper is not in the sense of Lahiri and Rao.

The rest of the paper is organized as follows. In Section 2, we discuss the model misspecification and rationale of the unconditional MSPE. We also derive some nice properties of the limit of the BPE and, as a consequence, the first-order unbiasedness of the PR MSPE estimator under the potential model misspecification. In Section 3, we propose a modification of the PR MSPE estimator that is second-order unbiased, again under the model misspecification, and provide a theoretical justification. In Section 4, we provide empirical evidence from simulation studies that fully supports our theoretical findings. A real-data example on the prediction of the incubation period of Covid-19 is discussed in Section 5 to further demonstrate the advantage of the new MSPE estimators. Proofs and further details are deferred to the Supplementary Material.

2. First-Order Unbiasedness of PR MSPE Estimator

In the SAE literature, different types of MSPEs have been considered, including conditional and unconditional ones; see, for example, Datta et al. (2011) for a review and discussion. In most cases, the MSPE is conditional on x, the covariates or auxiliary data; in other words, the auxiliary data are considered fixed. However, when model misspecification is taken into account, an MSPE conditional on x may not be reasonable, especially if one wishes to consider potential model misspecification when measuring uncertainty.

To see the rationale of the unconditional MSPE, it may be helpful to first understand that of the conditional MSPE. Note that when the underlying model is correctly specified, the small area mean, θ_i, is a linear function of x_i, plus an area-specific random effect, v_i [see above (1.2)]. Thus, knowing x_i has helped reduce

the uncertainty in θ_i. Therefore, when the underlying model is correctly specified, one can focus on the MSPE conditional on the auxiliary data, x_i, because there is no need to worry about uncertainty due to the latter. However, when the underlying model is misspecified, knowing x_i does not take away its uncertainty contribution to θ_i. In fact, because one is not sure how, or even whether, θ_i is associated with x_i, it may not even make sense to use the MSPE conditional on x_i as a measure of uncertainty.

There are other reasonable features of the unconditional MSPE. In practice, some of the auxiliary data are often sampled randomly with the responses, y, and are therefore subject to the same sampling variation. For example, in the hospital data of Morris and Christiansen (1995), which was used by Jiang, Nguyen and Rao (2011) to illustrate the OBP, the x_i denotes a severity index corresponding to the randomly selected hospitals. Suppose we wish to estimate or predict the mean failure rate of kidney transplants for hospital i, for $1 \leq i \leq m$, for the purpose of future planning. Obviously, for a future period of time, the x_i may not be the same as for the current period. Therefore, the potential future variation in x_i must be considered when assessing the uncertainty of the OBP.

For these reasons, we consider the MSPE with respect to the joint distribution of both the response, y, and the covariates, x. Specifically, we consider an area-level model, and assume that the true underlying model can be expressed as

$$y_i = x_i'\beta + \Delta_i + v_i + e_i, \tag{2.1}$$

for $i = 1, \ldots, m$. Here, Δ_i is an unobserved (random) variable satisfying the conditions below. Comparing (2.1) and (1.1), Δ_i represents the term of misspecification in the mean function. Let $z_i = (x_i', \Delta_i)'$ and $\delta_i = \Delta_i - x_i'\{\mathrm{E}(x_1x_1')\}^{-1}\mathrm{E}(x_1\Delta_1)$. We assume that the following assumptions hold:

A1. z_i, v_i, e_i, for $i = 1, \ldots, m$, are independent.

A2. $\mathrm{E}(z_i)$, $\mathrm{E}(z_iz_i')$, and $\mathrm{E}(\delta_i^4)$ are finite and do not depend on i, $v_i \sim N(0, A)$, and $e_i \sim N(0, D_i)$, where $A > 0$ is unknown and $D_i > 0$, for $1 \leq i \leq m$, is known.

An important feature of (2.1) is that it is not specific about the expression of Δ_i; thus, technically, (2.1) is no different from (1.4), which is the broader model used to derive the OBP (Jiang, Nguyen and Rao (2011)). Note that the true small area mean is $\theta_i = x_i'\beta + \Delta_i + v_i$, $i = 1, \ldots, m$. A difference between Δ_i and v_i is that v_i is assumed to have a specific distribution with mean zero and constant variance; no such assumption is made for Δ_i.

The standard LMM (e.g., Jiang (2007)) is a conditional model in that the covariates, x_i, are considered fixed. In contrast, the above model is considered a marginal model because x_i (and Δ_i) is a random variable. In the literature about correlated data, marginal models are known to be more robust to model misspecification when making inferences about characteristics of interest, such as the mean function (e.g., Song (2007)). However, we use the marginal model differently, because it is only used to evaluate a measure of uncertainty, that is, the MSPE. Here, we consider the unconditional, area-specific MSPE defined as

$$\mathrm{MSPE}(\hat{\theta}_i) = \mathrm{E}(\hat{\theta}_i - \theta_i)^2, \tag{2.2}$$

where the expectation is taken with respect to the joint distribution under model (2.1) with assumptions A1, A2.

Jiang, Nguyen and Rao (2011) derived an estimator of the MSPE conditional on x_i, and showed that the estimator is second-order unbiased. This is referred to as the JNR estimator hereafter. Note that, if $\widetilde{\mathrm{MSPE}}$ is a second-order unbiased estimator of the conditional MSPE, that is, $\mathrm{E}\{(\hat{\theta}_i - \theta_i)^2 | X\}$, then, under regularity conditions, $\widetilde{\mathrm{MSPE}}$ is also a second-order unbiased estimator of the unconditional MSPE, because $\mathrm{E}(\widetilde{\mathrm{MSPE}}) = \mathrm{E}\{\mathrm{E}(\widetilde{\mathrm{MSPE}}|X)\} = \mathrm{E}[\mathrm{E}\{(\hat{\theta}_i - \theta_i)^2|X\} + o_{\mathrm{P}}(m^{-1})] = \mathrm{MSPE} + o(m^{-1})$, applying regularity conditions to ensure that $\mathrm{E}\{o_{\mathrm{P}}(m^{-1})\} = o(m^{-1})$. A problem with the JNR estimator is that its leading $[O(1)]$ term involves a single observation, y_i, that has large variation. Specifically, the leading term of the JNR estimator is

$$(\hat{\theta}_i - y_i)^2 + D_i(2\hat{B}_i - 1), \tag{2.3}$$

where $\hat{B}_i$ is B_i with A replaced by $\hat{A}$. Owing to the involvement of y_i in the leading term, the JNR estimator has large variation; in fact, the variance of the JNR estimator is $O(1)$. Thus, there is a non-vanishing probability, even as the sample size increases, that the JNR MSPE estimator is negative. This can be seen because the second term in (2.3) becomes negative when $\hat{B}_i < 1/2$. For example, Jiang, Nguyen and Rao (2011) computed the JNR estimator for the 23 hospitals in the kidney transplant data of Morris and Christiansen (1995), finding that the value of the JNR estimator is negative for six out of the 23 hospitals.

In contrast, the PR MSPE estimator is stable in the sense that it is a function of $\hat{A}$, an estimator based on data from all of the small areas. For example, the leading term of the PR estimator is given by (e.g., Datta and Lahiri (2000))

$$\frac{\hat{A}D_i}{(\hat{A} + D_i)}. \tag{2.4}$$

Here, (2.4) is typically denoted by $g_{1i}(\hat{A})$, following the notation of Prasad and Rao (1990). Note that $g_{1i}(A)$ is the MSPE of the BP when the underlying model is correctly specified, and A is the true variance of v_i. It is easy to show that, under regularity conditions, the variance of (2.4) is $O(m^{-1})$. As a result, the variance of the PR estimator is also $O(m^{-1})$. Another well-known feature of the PR MSPE estimator is that it is second-order unbiased, provided that the underlying model is correctly specified. However, here, the underlying model is misspecified in terms of the mean function. Could it be that, somehow, the PR MSPE estimator is still correct, or mostly correct, in spite of the model misspecification?

The short answer is yes. To see why this is even possible, note that, according to the theory established in Jiang, Nguyen and Rao (2011), the BPE of $\psi = (\beta', A)'$, denoted by $\hat{\psi} = (\hat{\beta}, \hat{A})'$, converges in probability to the minimizer of

$$q(\psi) = \mathrm{E}\{|\tilde{\theta}(\psi) - \theta|^2\} = \sum_{i=1}^{m} \mathrm{E}\{\tilde{\theta}_i(\psi) - \theta_i\}^2, \tag{2.5}$$

where $\tilde{\theta}(\psi) = [\tilde{\theta}_i(\psi)]_{1\le i\le m}$ and $\tilde{\theta}_i(\psi)$ is the right side of (1.2) when ψ is treated as a parameter vector (rather than the true parameter vector). Here, the expectation is taken with respect to the true underlying distribution. Let $\psi_* = (\beta_*', A_*)'$ denote the minimizer of (2.5). Note that ψ_* is not necessarily equal to the true parameter vector under (2.1), denoted by $\psi_0 = (\beta_0', A_0)'$, if the underlying model is misspecified (i.e., $\Delta_i \neq 0$). If one takes the limit of (2.4) by letting $m \to \infty$, the leading term of the PR estimator becomes

$$\frac{A_* D_i}{(A_* + D_i)}. \tag{2.6}$$

To obtain an expression for ψ_*, note that one can derive the expression

$$\begin{aligned} \mathrm{E}\{\tilde{\theta}_i(\psi) - \theta_i\}^2 &= \mathrm{E}\left(\frac{A}{A+D_i} y_i + \frac{D_i}{A+D_i} x_i'\beta - \theta_i\right)^2 \\ &= \left(\frac{A}{A+D_i}\right)^2 D_i + \left(\frac{D_i}{A+D_i}\right)^2 [A_0 + \mathrm{E}\{x_1'(\beta - \beta_0) - \Delta_1\}^2] \end{aligned} \tag{2.7}$$

using assumptions A1, A2 above, noting that the true small area mean is $\theta_i = x_i'\beta_0 + \Delta_i + v_i$ with $v_i \sim N(0, A_0)$. By differentiating q in (2.5) with respect to β and A and letting the derivatives equal to zero, we can derive the following:

$$\beta_* = \beta_0 + \{\mathrm{E}(x_1 x_1')\}^{-1} \mathrm{E}(x_1 \Delta_1), \tag{2.8}$$

$$A_* = A_0 + \mathrm{E}\{x_1'(\beta_* - \beta_0) - \Delta_1\}^2. \tag{2.9}$$

A consequence of (2.8) and (2.9) is that, under the misspecified model, we have

$$\mathrm{E}\{x_i(y_i - x_i'\beta_*)\} = 0, \quad \mathrm{E}(y_i - x_i'\beta_*)^2 = A_* + D_i. \tag{2.10}$$

A remarkable feature of ψ_* is that, although it is the minimizer of the overall MSPE, (2.5), which is the sum of the MSPEs for different small areas, it actually minimizes the MSPE for all small areas. We formally state the result as follows.

Theorem 1. *The parameter vector ψ_*, defined as the minimizer of* (2.5) *satisfying* (2.8) *and* (2.9)*, minimizes* $\mathrm{E}\{\tilde{\theta}_i(\psi) - \theta_i\}^2$ *over all ψ, for every $1 \le i \le m$.*

Proof. Write $\tilde{\theta}_i(\psi) = g_i(\psi, y_i)$ and note that $\theta_i = y_i - e_i$. It can then be shown, using (2.10) in particular (see the Supplementary Material) that for any ψ, we have

$$\begin{aligned}\mathrm{E}\{g_i(\psi, y_i) - \theta_i\}^2 &= \mathrm{E}\{g_i(\psi_*, y_i) - \theta_i\}^2 + \mathrm{E}\{g_i(\psi, y_i) - g_i(\psi_*, y_i)\}^2 \\ &\ge \mathrm{E}\{g_i(\psi_*, y_i) - \theta_i\}^2,\end{aligned} \tag{2.11}$$

with the equality on the right side of (2.11) holding if and only if $\mathrm{E}\{g_i(\psi, y_i) - g_i(\psi_*, y_i\}^2 = 0$. This argument applies to any $1 \le i \le m$.

On the other hand, the true MSPE of $\hat{\theta}_i$, by (2.2), is

$$\mathrm{MSPE}(\hat{\theta}_i) = \mathrm{E}\{g_i(\hat{\psi}, y_i) - \theta_i\}^2, \tag{2.12}$$

where $\hat{\beta}$ and $\hat{A}$ are the BPEs, and θ_i is the true small area mean. Because $\hat{\beta} \xrightarrow{\mathrm{P}} \beta_*, \hat{A} \xrightarrow{\mathrm{P}} A_*$, under regularity conditions, the leading term of the true MSPE is (2.12), with $\hat{\beta}$ and $\hat{A}$ replaced by β_* and A_*, respectively, which is equal to the right side of (2.7) with β and A replaced by β_* and A_*, respectively. Therefore, by (2.7), (2.9), and (2.6), we conclude the following:

$$\begin{aligned}&\text{Leading term of the true MSPE} \\ &= \left(\frac{A_*}{A_* + D_i}\right)^2 D_i + \left(\frac{D_i}{A_* + D_i}\right)^2 [A_0 + \mathrm{E}\{x_1'(\beta_* - \beta_0) - \Delta_1\}^2] \\ &= \left(\frac{A_*}{A_* + D_i}\right)^2 D_i + \left(\frac{D_i}{A_* + D_i}\right)^2 A_* = \frac{A_* D_i}{A_* + D_i} \\ &= \text{Leading term of the PR MSPE estimator.}\end{aligned} \tag{2.13}$$

Note that, typically, the leading terms are $O(1)$, while the remaining terms are $O(m^{-1})$. Thus, (2.13) makes us believe that the PR estimator is, at least,

first-order unbiased. The latter conjecture is, indeed, correct (which is implied by the arguments of the proof of Theorem 2).

However, as it turns out, under the current model misspecification, the PR estimator is not "second-order robust" as in Lahiri and Rao (1995); see the note in Section 1). One reason for this is that the OBP is not the BP, even with $\hat{\psi}$ replaced by ψ_*. Recall that ψ_* is not the true parameter vector. In fact, with the form of Δ_i in (2.1) unknown, it is not even clear what the true parameter vector is under the assumed model, (1.1). As a result, one does not have the well-known orthogonal decomposition of the MSPE that is associated with the BP. Specifically, one has

$$\begin{aligned}\mathrm{MSPE}(\hat{\theta}_i) = c_i(A_*) + 2\mathrm{E}[\{g_i(\psi_*, y_i) - \theta_i\}\{g_i(\hat{\psi}, y_i) - g_i(\psi_*, y_i)\}] \\ +\mathrm{E}\{g_i(\hat{\psi}, y_i) - g_i(\psi_*, y_i)\}^2,\end{aligned} \tag{2.14}$$

where $c_i(A) = \mathrm{E}\{g_i(\psi_*, y_i) - \theta_i\}^2 = AD_i/(A + D_i)$. Under a correct specification of the model, $\psi_* = \psi_0$, and hence $g_i(\psi_*, y_i)$ is the BP. As a result, the second term on the right side of (2.14) vanishes, and one gets the orthogonal decomposition. However, this (typically) is no longer true under the misspecified model (2.1). In fact, it is shown (see the proof of Theorem 2 in the Supplementary Material) that the second term on the right side of (2.14) is $O(m^{-1})$ under the misspecified model. Because the PR estimator only provides a second-order unbiased estimation of the first and third terms on the right side of (2.14), assuming that the orthogonal decomposition holds, it follows that the PR estimator is not second-order unbiased for estimating the MSPE of the OBP under the model misspecification. There are also other reasons why the PR estimator is not second-order unbiased (see the next section for further explanation).

3. A Modified PR MSPE Estimator

Nevertheless, the PR estimator can be modified to achieve second-order unbiasedness. There are two issues that we need to handle in order to come up with a second-order unbiased modification of the PR estimator, called the MPR estimator. First, as noted earlier, the middle term on the right side of (2.14) is $O(m^{-1})$, rather than zero or even $o(m^{-1})$. This term needs to be estimated with a bias of $o(m^{-1})$.

Second, although the parameter vector ψ_* possesses some nice properties, such as (2.10), these only work up to the second moments. For example, the fourth moment of $y_i - x_i'\beta_*$ does not follow that of a zero-mean normal distribution. Note that if $\Delta_i = 0$, then $\beta_* = \beta_0$, and hence $y_i - x_i'\beta_* = v_i + e_i \sim$

$N(0, A + D_i)$. Thus, we have $\mathrm{E}(y_i - x_i'\beta_*)^4 = 3(A + D_i)^2$. However, the latter may not hold when $\Delta_i \neq 0$. When this difference is ignored, a bias of order $O(m^{-1})$ is induced. Thus, to achieve second-order unbiasedness, a solution must be found to the difficulties related to the higher moments. Here, we use a similar idea to the observed information (e.g., Efron and Hinkley (1978)). For example, if $\mathrm{E}(y_i - x_i'\beta_*)^4$ cannot be expressed as a known function of ψ_*, we do not attempt to compute the expectation. Instead, we remove the expectation sign, and replace the expression inside by its estimated version, that is, $(y_i - x_i'\hat{\beta})^4$. In this example, the replacement does not yield a second-order unbiased estimator, but this is not what we are actually dealing with. Luckily, we only have to make such replacements when the term is already $O(m^{-1})$. In this case, the replacement results in a difference of lower order, that is, $o(m^{-1})$, which is what we are looking for.

Another point to keep in mind is that we need to avoid using a term that involves a single observation, such as (2.3) in the JNR estimator, in our MSPE estimator. This is because such a term has high variation, even if second-order unbiasedness can be achieved by using such a term. Thus, for example, a single term such as $(y_i - x_i'\hat{\beta})^4$ is replaced by a weighted average of such terms over all of the small areas in order to reduced the variation.

The MPR estimator is given below. By Jiang, Nguyen and Rao (2011), the BPE of ψ, $\hat{\psi} = (\hat{\beta}', \hat{A})'$ is the minimizer of $Q(\psi, y) = \sum_{j=1}^m Q_j(\psi, y_j)$, with

$$Q_j(\psi, y_j) = \left(\frac{D_j}{A + D_j}\right)^2 (y_j - x_j'\beta)^2 + \frac{2AD_j}{A + D_j}. \tag{3.1}$$

Let $p = \dim(\beta)$. Define $\hat{r}_i = D_i/(\hat{A} + D_i)$, $\hat{t}_m = \sum_{j=1}^m (\hat{A} + D_j)^{-2}$,

$$\hat{s}_{km} = \sum_{j=1}^m \frac{\hat{r}_j^2}{(\hat{A} + D_j)^k},\ k = 0, 1, 2, \quad \hat{u}_{km} = p\sum_{j=1}^m \frac{D_j \hat{r}_j^3}{(\hat{A} + D_j)^k},\ k = 0, 1.$$

Next, define the following quantities:

$$\hat{T}_m = \sum_{j=1}^m \left\{\frac{(y_j - x_j'\hat{\beta})^4}{(\hat{A} + D_j)^2} - 3\right\}, \quad \hat{V}_{km} = \sum_{j=1}^m \frac{\hat{r}_j^4}{(\hat{A} + D_j)^k}\left\{\frac{(y_j - x_j'\hat{\beta})^4}{(\hat{A} + D_j)^2} - 1\right\},$$

for $k = 0, 1$. Finally, define $\hat{a}_i = \hat{r}_i^4 \hat{T}_m/(\hat{A} + D_i)^2 \hat{s}_{1m}\hat{t}_m$ and

$$\hat{b}_i = \hat{r}_i^2 \left\{\frac{2\hat{u}_{1m}}{\hat{s}_{0m}\hat{s}_{1m}} + \frac{3}{\hat{s}_{1m}^3}(\hat{s}_{1m}\hat{V}_{1m} - \hat{s}_{2m}\hat{V}_{0m}) + \frac{2\hat{V}_{0m}}{(\hat{A} + D_i)\hat{s}_{1m}^2}\right\}.$$

The MPR estimator of $\mathrm{MSPE}(\hat{\theta}_i)$ is given by

$$\widehat{\mathrm{MSPE}}(\hat{\theta}_i) = c_i(\hat{A}) - 2\hat{a}_i + \hat{b}_i, \tag{3.2}$$

where $c_i(\cdot)$ is defined below (2.14). Interpreting separately, $c_i(\hat{A}) + \hat{b}_i$ is the combination of a bias-correction term to $c_i(\hat{A})$ and an estimator of the last term on the right side of (2.14); furthermore, $-\hat{a}_i$ is an estimator of the expected cross-product on the right side of (2.14).

The PR MSPE estimator, without modification, is derived in the same way as the MPR estimator, except we compute the fourth moments of $y_i - x_i'\beta_*$ under the assumption that the underlying model is correct. Under the latter assumption, we have $\mathrm{E}(y_i - x_i'\beta_*)^4 = 3(A_* + D_i)^2$ (note that $\psi_* = \psi_0$ when there is no model misspecification). Therefore, the PR estimator is given by the same expression of the MPR estimator, except with $\hat{a}_i$ replaced by zero, and $\hat{V}_{km}$ replaced by $\hat{v}_{km} = 2\sum_{j=1}^m \hat{r}_j^4/(\hat{A} + D_j)^k$, for $k = 0, 1$.

To justify the second-order unbiasedness of the proposed MPR estimator, we follow the treatment of Das, Jiang and Rao (2004) (also see Jiang, Lahiri and Wan (2002)) to regularize the BPE, as follows. Let $\hat{\psi}$ be unchanged if $|\hat{\psi}| \leq C\{\log(m+1)\}^K$, where C and K are any given positive constants; otherwise, let $\hat{\psi} = \psi_{\mathrm{f}}$, where ψ_{f} is any known vector satisfying $|\psi_{\mathrm{f}}| \leq C\{\log(m+1)\}^K$. Note that such a truncation make no difference in practice, because one can argue that there are always constants C and K such that $|\hat{\psi}| \leq C\{\log(m+1)\}^K$ holds for the data. Therefore, there is no need to truncate $\hat{\psi}$ in practice; the truncation is only for the theoretical argument below. In addition, following Jiang, Lahiri and Wan (2002), the estimator $\hat{\psi}$ is said to be consistent uniformly (c.u.) at rate m^{-d} if, for any $\delta > 0$, there is a constant c_δ such that $\mathrm{P}(\mathcal{A}_\delta^c) \leq c_\delta m^{-d}$, where $\mathcal{A}_\delta = \{\partial\hat{Q}/\partial\psi = 0, |\hat{\psi} - \psi_*| \leq \delta\}$, with $\partial\hat{Q}/\partial\psi = \partial Q/\partial\psi|_{\hat{\psi}}$. In addition to assumptions A1, A2 of Section 2, we assume the following regularity conditions:

A3. $\mathrm{E}(|z_j|^{10})$, for $1 \leq j \leq m$, are bounded, $\mathrm{E}(x_1x_1')$ is nonsingular, and there are constants $0 < b \leq B < \infty$ such that $b \leq D_i \leq B$, for $1 \leq i \leq m$.

A4. $\hat{\psi} - \psi_* = O_{\mathrm{P}}(m^{-1/2})$, and $\hat{\psi}$ is c.u. at rate d, for some $d > 2$.

The nonsingularity of $\mathrm{E}(x_1x_1')$ in assumption A3 is needed for expression (2.8) to hold. Jiang, Nguyen and Rao (2011) give sufficient conditions for the root-m consistency of $\hat{\psi}$, which is the first part of assumption A4. For the verification of the c.u. condition, see Jiang, Lahiri and Wan (2002). The second-order unbiasedness of the MPR estimator is established by the following theorem.

Theorem 2. *Under assumptions A1–A4, the MPR estimator,* (3.2)*, is second-order unbiased; that is,* $\mathrm{E}\{\widehat{\mathrm{MSPE}}(\hat{\theta}_i)\} = \mathrm{MSPE}(\hat{\theta}_i) + o(m^{-1})$ *for every* i.

Outline of the proof: Where are the modifications? The proof of Theorem 2 is given in the Supplementary Material. An outline of the proof is given below, which also explains where the modifications of the PR estimator take place. The first modification deals with the middle term on the right side of (2.14), which is $O(m^{-1})$. The PR estimator does not involve this term (because it is zero when the model is correctly specified). Denote the middle term by $2I_1$. It is shown that $I_1 = -\mathrm{E}(\hat{a}_i) + o(m^{-1})$, where $\hat{a}_i$ is given above (3.2). Two important identities, which can be derived from (2.10), are used to simplify the expression of $\mathrm{E}(\hat{a}_i)$ and obtain its second-order unbiased estimator, namely,

$$\mathrm{E}[x_i'\{\mathrm{E}(x_1x_1')\}^{-1}x_i(y_i - x_i'\beta_*)^2] = (A_* + D_i)p; \tag{3.3}$$

$$\frac{\mathrm{E}(y_i - x_i'\beta_*)^4}{(A_* + D_i)^2} - 3 = \frac{\mathrm{K}(\delta_1)}{(A_* + D_i)^2}, \tag{3.4}$$

where δ_i is defined below (2.1), and the kurtosis of a random variable, ξ, is defined as $\mathrm{K}(\xi) = \mathrm{E}(\xi^4) - 3\{\mathrm{E}(\xi^2)\}^2$. Note that (3.4) holds for every $1 \le i \le m$. This leads to a way to "solve" for the unknown $\mathrm{K}(\delta_1)$ by taking summations of both sides of the equation over i, leading to

$$\mathrm{K}(\delta_1) = \mathrm{E}\left[\left\{\sum_{j=1}^{m} \frac{1}{(A_* + D_j)^2}\right\}^{-1} \sum_{j=1}^{m} \left\{\frac{(y_j - x_j'\beta_*)^4}{(A_* + D_j)^2} - 3\right\}\right]. \tag{3.5}$$

In fact, expression (3.5) helps to obtain a bias-corrected estimator of I_1.

The second modification has to do with the last term on the right side of (2.14), denoted by I_2. This term is known to have the following kind of expression (e.g., Prasad and Rao (1990), Datta and Lahiri (2000)):

$$I_2 = \sum_{j=1}^{m} \mathrm{E}(u_j'G^{-1}R_iG^{-1}u_j) + o(m^{-1}), \tag{3.6}$$

where $u_j = \partial Q_j/\partial\psi|_{\psi_*}$, $G = \sum_{j=1}^{m} \mathrm{E}(\partial^2 Q_j/\partial\psi\partial\psi'|_{\psi_*})$, and R_i is another matrix that depends on A_* and $\mathrm{E}(x_1x_1')$. Expressions of u_j, G, and R_i can be obtained. The expression of the first term on the right side of (3.6) can be simplified, with the help of (2.10) and (3.3); however, it still involves $\mathrm{E}(y_j - x_j'\beta_*)^4$, for $1 \le j \le m$. Prasad and Rao computed these fourth moments under the correct specification of the model, that is, $\mathrm{E}(y_j - x_j'\beta_*)^4 = 3(A_* + D_j)^2$. In our case, the latter identity does not hold, owing to the model misspecification, and the fourth moments are

are not (known) functions of ψ_*. When the Prasad-Rao calculation is mistakenly used, it results in a bias of the order $O(m^{-1})$. In our modification, the first term on the right side of (3.6) is estimated in a way similar to the observed information noted in the second paragraph of this section. This way, we can express I_2 as $\mathrm{E}(\hat{d}_i) + o(m^{-1})$, where $\hat{d}_i$ is part of $\hat{b}_i$ in (3.2).

The final modification has to do with the bias correction to $c_i(\hat{A})$. Here, we once again encounter the issues related to the fourth moments. The same strategy is used to solve the problem. This leads to the expression $c_i(A_*) = \mathrm{E}\{c_i(\hat{A}) + \hat{h}_i\} + o(m^{-1})$, where $\hat{h}_i$ is another part of $\hat{b}_i$ in (3.2).

4. Simulation Studies

In this section, we carry out simulation studies to investigate the finite-sample performance of different estimators of the MSPE of the OBP under various scenarios, where the underlying model is correctly specified or misspecified. As noted, the OBP may be viewed as an EBP in which the model parameters are estimated by the BPE (see Section 1). Specifically, we compare the two proposed PR MSPE estimators for the OBP proposed in the previous sections, and their comparison with the existing MSPE estimators for the OBP, including the JNR estimator (see Section 2) and a naive MSPE estimator for OBP, also considered by Jiang, Nguyen and Rao (2011) for comparison purposes. The computational details of the simulations are given in the Supplementary Material in algorithmic form.

Three examples are considered to account for different kinds of scenarios, with or without model misspecification. In all examples, the assumed model can be expressed as $y_i = \beta_0 + x_i'\beta + v_i + e_i$, for $i = 1, 2, \ldots, m$, and the true model takes the following forms, respectively, for the three examples:

(I). $y_i = \beta_0 + x_i'\beta + v_i + e_i$, for $i = 1, 2, \ldots, m$;

(II). $y_i = \beta_0 + \beta_1 x_{1i} + \arctan(z_i) + v_i + e_i$, for $i = 1, 2, \ldots, m$;

(III). $y_i = \arctan(4x_i'\beta) + v_i + e_i$, for $i = 1, 2, \ldots, m$.

Here, $\beta_0 = 0.2$, $\beta_1 = 0.5$, and $x_i = (x_{1i}, x_{2i})'$ such that $(x_i', z_i)'$ has a trivariate normal distribution with means zero, variances two, $\mathrm{cov}(x_{1i}, x_{2i}) = \mathrm{cov}(x_{2i}, z_i) = 0.4$, and $\mathrm{cov}(x_{1i}, z_i) = 0$; $v_i \sim N(0, A)$ with $A = 1$, and $e_i \sim N(0, D_i)$ with $D_i = 0.5 + (i-1)/(m-1)$, for $i = 1, 2, \ldots, m$. In Example (I), the assumed model is the same as the true one; hence, there is no model misspecification. In Example (II), the nonlinear part of the true model is misspecified as linear, indicating that

Table 1. Comparison of MSPE Estimators in Mean %RB, Mean %|RB|, and %NE.

	m	MPR		PR		Naive			JNR		
		M%RB	M%|RB|	M%RB	M%|RB|	M%RB	M%|RB|	%NE	M%RB	M%|RB|	%NE
(I)	20	4.55	4.69	4.85	5.00	−47.44	47.44	36	4.57	5.00	21
	40	0.25	1.30	0.66	1.45	−25.28	25.28	27	0.31	1.61	19
	80	−0.05	1.11	0.12	1.12	−13.03	13.03	22	0.01	1.80	18
	160	−0.02	1.13	0.02	1.13	−6.59	6.61	19	−0.00	1.61	16
	320	0.03	1.12	0.04	1.12	−3.28	3.42	17	0.03	1.64	16
	640	−0.03	1.15	−0.04	1.15	−1.70	2.26	16	−0.03	1.70	16
(II)	20	−0.09	1.17	1.99	1.99	−33.37	33.38	19	0.93	1.44	9
	40	−0.90	1.38	−0.07	1.12	−17.29	17.30	10	−0.36	1.53	5
	80	−0.06	1.15	0.19	1.15	−8.32	8.32	4	0.11	1.43	3
	160	−0.16	1.16	−0.09	1.15	−4.32	4.42	2	−0.11	1.27	1
	320	0.14	1.18	0.15	1.18	−1.95	2.24	1	0.15	1.45	< 1
	640	0.04	1.14	0.04	1.14	−1.01	1.65	< 1	0.04	1.37	< 1
(III)	20	0.41	1.70	1.83	2.26	−39.20	39.21	26	0.38	1.80	13
	40	−0.10	1.19	0.64	1.37	−19.50	19.51	16	0.33	1.38	10
	80	−0.11	1.19	0.12	1.21	−9.93	9.93	10	0.02	1.42	7
	160	−0.10	1.20	−0.03	1.20	−5.04	5.07	6	−0.04	1.54	5
	320	0.15	1.11	0.17	1.12	−2.33	2.60	4	0.17	1.48	4
	640	0.02	1.11	0.01	1.11	−1.23	1.80	3	0.02	1.38	3

the assumed model is partly misspecified. In Example (III), all predictors are nonlinearly associated with the response; hence, the model is misspecified.

For each example, we carried out 10,000 simulation runs. The sample size, m, varies from 20 to 640. The MSPE estimators, MPR, PR, Naive, and JNR, are first compared in terms of their percentage relative bias, defined as

$$\%\mathrm{RB} = 100 \times \left[\frac{\mathrm{E}(\widehat{\mathrm{MSPE}}) - \mathrm{MSPE}}{\mathrm{MSPE}}\right],$$

where $\widehat{\mathrm{MSPE}}$ denotes an MSPE estimator, $\mathrm{E}(\widehat{\mathrm{MSPE}})$ is the simulated mean of $\widehat{\mathrm{MSPE}}$, and MSPE is the true MSPE, which is evaluated empirically (see Section A.3 of the Supplementary Material). The %RB is then averaged over different small areas, and the mean %RB (M%RB) is reported in Table 1. Because the M%RB involves the average of %RB over different small areas, positive and negative %RBs may cancel each other. To take this into account, we also report the mean absolute value of the %RBs over the small areas (M%|RB|, i.e., taking the absolute value before averaging over the small areas). The standard deviations corresponding to the M%RB and %|RB| are deferred to the Supplementary Material.

Box plots of the %RB, corresponding to $m = 20, 40, 80$, are presented in Figure 1 for the %RBs for different small areas. Note that the box plots provide

information about the %RB for individual small areas, not just the average %RB. Therefore, the pictures are quite clear even without the box plots for the absolute values of the %RBs.

Also reported in Table 1 are the percentage of negative MSPE estimates (%NE). The PR and MPR estimates were positive in all simulation runs, but some of the naive and JNR estimates were negative. The %NE, observed in the simulation runs and averaged over the small areas, is reported for the naive MSPE estimator and JNR.

It is clear that MPR, PR, and JNR significantly outperform the naive MSPE estimator in terms of both %RB and %|RB|, which is not surprising. Both MPR and JNR are second-order unbiased, while the naive MSPE estimator is only first-order unbiased (Jiang, Nguyen and Rao (2011)). Although PR is also first-order unbiased, its performance is superb, at least for this simulation study. In fact, it seems that MPR and PR perform better in terms of the standard deviations of the %RB and %|RB| (over the small areas) than does JNR (see Supplementary Material). In terms of M%RB, MPR and JNR seem to perform slightly better than PR when m is relatively small; when m is relatively large, MPR, PR, and JNR are quite close in terms of M%RB. In terms of M%|RB|, MPR outperforms JNR in all cases; in most cases, MPR outperforms PR, or the two perform equally well. As noted [see the paragraph below (3.2)], the only difference between the PR and MPR estimators is the treatment of

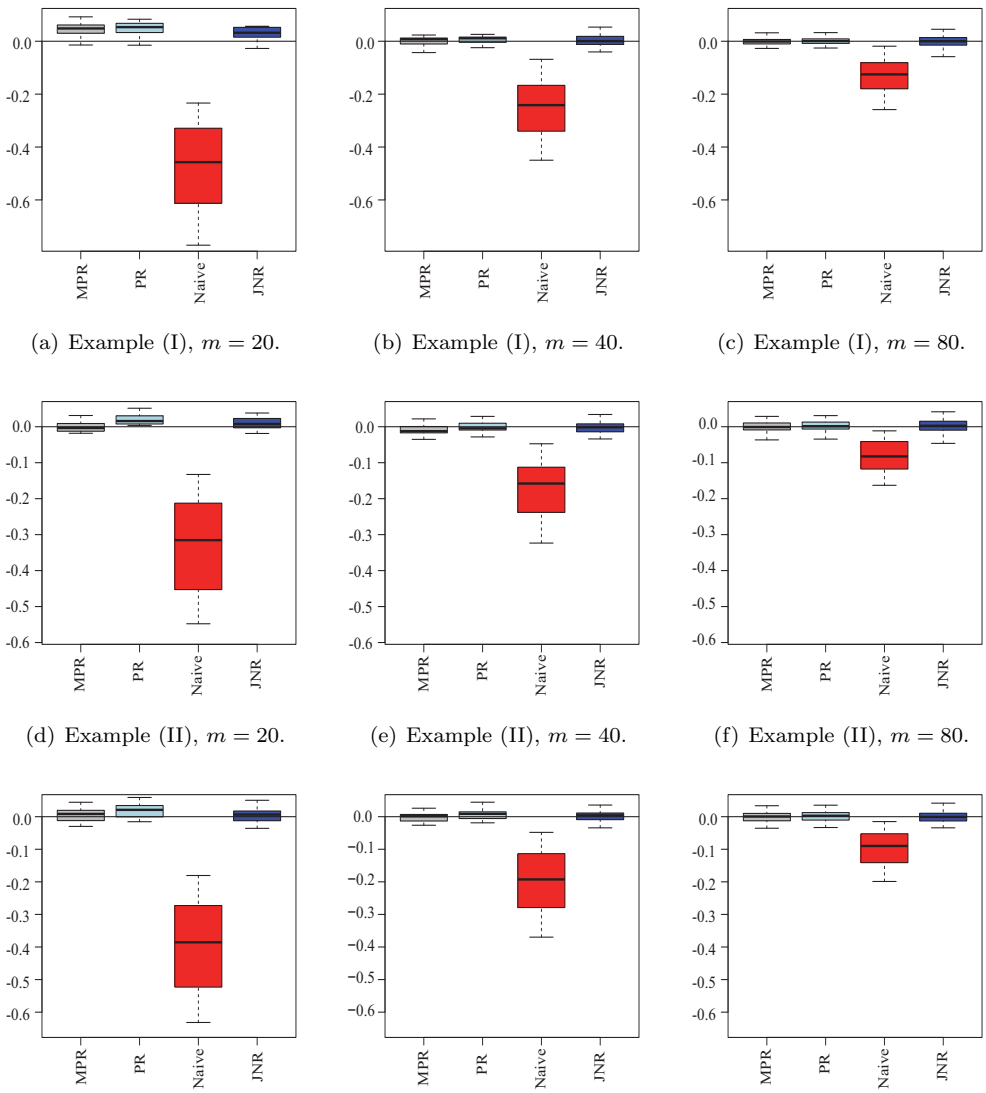
$$\mathrm{E}(y_j - x_j'\beta_*)^4, \quad 1 \le j \le m \tag{4.1}$$

in the estimation. For PR, the terms in (4.1) are replaced by

$$3(\hat{A} + D_j)^2, \quad 1 \le j \le m, \tag{4.2}$$

leading to simplified expressions. For MPR, (4.1) are estimated using the observed information approach. Although, theoretically speaking, MPR is second-order unbiased while PR is first-order unbiased, at least in this simulation study, the performance of MPR and PR are quite close, especially for larger m.

Although, in terms of %RB, JNR appears to be doing well, the real difference between JNR and the two PR estimators is in the variance. Table 1 shows that the NE% of JNR can be as high as 20%, indicating a significant chance that the JNR estimator is negative; for the naive MSPE estimator, the %NE can be as high as 35%. In contrast, the two PR estimators never take negative values in the simulation runs. Note that a negative value for an MSPE estimator is unpleasant, and hence should be avoided as much as possible.



(a) Example (I), $m = 20$. (b) Example (I), $m = 40$. (c) Example (I), $m = 80$.

(d) Example (II), $m = 20$. (e) Example (II), $m = 40$. (f) Example (II), $m = 80$.

(g) Example (III), $m = 20$. (h) Example (III), $m = 40$. (i) Example (III), $m = 80$.

Figure 1. Box plots of %RB for Different MSPE Estimators.

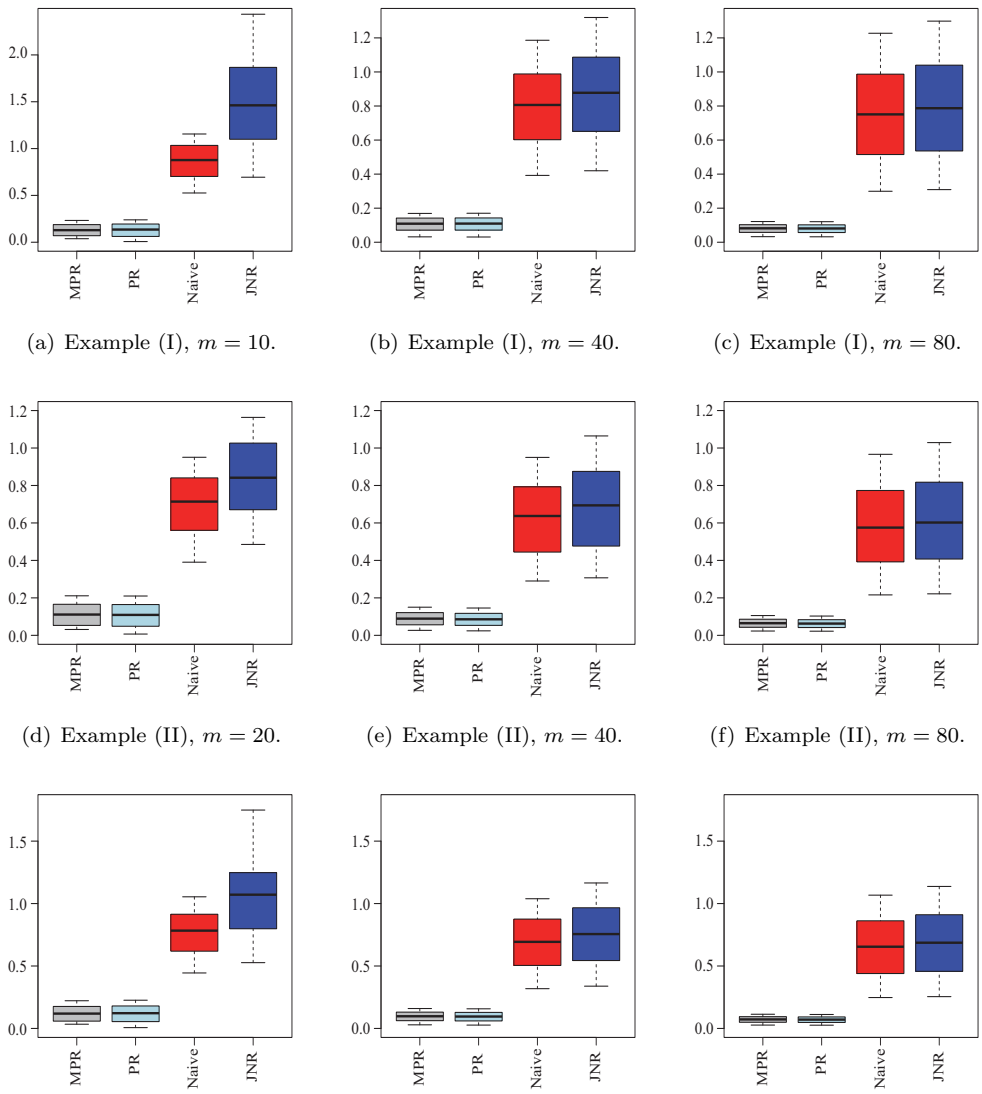
Table 2 and Figure 2 show the performance of the MSPE estimators in terms of the variance. Presented in Table 2 are summaries of the simulated standard deviations (s.d.) of the MSPE estimators. For each small area, we obtain the s.d. of the MSPE estimator, and then report the mean and standard deviation of the simulated s.d. under each scenario and sample size, and for each method. Figure 2 presents box plots of the simulated s.d.'s. It is clear from Table 2 and Figure

Table 2. Mean s.d. of MSPE Estimators (the Number in Parentheses Show the Ratio of the Mean s.d. of the Corresponding Estimator to That of the MPR Estimator).

Example	m	MPR	PR	Naive	JNR
(I)	20	0.1295 (1.00)	0.1286 (0.99)	0.8616 (6.65)	1.5666 (12.10)
	40	0.1058 (1.00)	0.1063 (1.00)	0.7992 (7.55)	0.8741 (8.26)
	80	0.0801 (1.00)	0.0793 (0.99)	0.7605 (9.49)	0.7984 (9.96)
	160	0.0578 (1.00)	0.0573 (0.99)	0.7427 (12.85)	0.7618 (13.19)
	320	0.0408 (1.00)	0.0407 (1.00)	0.7304 (17.88)	0.7399 (18.11)
	640	0.0292 (1.00)	0.0291 (1.00)	0.7269 (24.88)	0.7318 (25.05)
(II)	20	0.1141 (1.00)	0.1077 (0.94)	0.7015 (6.15)	0.8492 (7.44)
	40	0.0898 (1.00)	0.0864 (0.96)	0.6245 (6.95)	0.6833 (7.61)
	80	0.0634 (1.00)	0.0616 (0.97)	0.5842 (9.21)	0.6142 (9.68)
	160	0.0455 (1.00)	0.0448 (0.98)	0.5659 (12.43)	0.5810 (12.77)
	320	0.0319 (1.00)	0.0316 (0.99)	0.5573 (17.49)	0.5649 (17.73)
	640	0.0229 (1.00)	0.0228 (1.00)	0.5527 (24.13)	0.5565 (24.30)
(III)	20	0.1294 (1.00)	0.1278 (0.99)	0.8606 (6.65)	2.4415 (18.87)
	40	0.1070 (1.00)	0.1072 (1.00)	0.8011 (7.48)	0.8834 (8.25)
	80	0.0802 (1.00)	0.0793 (0.99)	0.7598 (9.47)	0.7976 (9.94)
	160	0.0583 (1.00)	0.0578 (0.99)	0.7421 (12.74)	0.7612 (13.06)
	320	0.0410 (1.00)	0.0408 (1.00)	0.7310 (17.84)	0.7406 (18.08)
	640	0.0291 (1.00)	0.0291 (1.00)	0.7273 (24.95)	0.7321 (25.12)

2 that the two PR methods have significant advantage over the naive MSPE estimator and JNR in terms of the variance. Between the two PR estimators, it appears that the mean s.d. of PR is slightly smaller than that of MPR, which is reasonable. As noted, MPR uses the observed information approach, which results in a slight increase of the variance of the MSPE estimator. The numbers in parentheses report the ratio of the mean s.d. of the corresponding MSPE estimator to that of the MPR estimator. It is seen that the ratios for PR are either equal to one or slightly less than one, while the ratios for Naive and JNR are all much larger than one (some as high as 25). It is also seen (especially from Figure 2) that JNR performs worse than Naive in term of the variance.

In conclusion, the simulation results show that the two PR estimators perform extremely well, both in terms of the bias and in terms of the variation. Especially in terms of the variation, the two PR estimators significantly outperform the naive and JNR estimators. In addition, unlike the naive and JNR estimators, no negative values are observed for the two PR estimators. All three estimators, MPR, PR, and JNR, significantly outperform the naive MSPE estimator in terms of the bias. Between the two PR estimators, it appears that MPR performs slightly better than PR in terms of the bias, and PR performs slightly



(a) Example (I), $m = 10$. (b) Example (I), $m = 40$. (c) Example (I), $m = 80$.

(d) Example (II), $m = 20$. (e) Example (II), $m = 40$. (f) Example (II), $m = 80$.

(g) Example (III), $m = 20$. (h) Example (III), $m = 40$. (i) Example (III), $m = 80$.

Figure 2. Box plots of Simulated s.d. for Different MSPE Estimators.

better than MPR in terms of the variation.

5. Applications: Prediction of Incubation Period for Covid-19

Since the first case reported in Wuhan, China in December 2019, the 2019 coronavirus disease (Covid-19) there has been a rapid increase in cases and deaths.

From a scientific point of view, the incubation period (IP) of the disease plays an important role in prevention and control efforts, as well as mathematical modeling of the coronavirus transmission. The current quarantine period in China is fixed at 14 days, without considering any other auxiliary information; see, for example, Guan et al. (2020), Li et al. (2020) and Linton et al. (2020).

A complication of the existing IP in the human population is that it may have to do with both the nature of the disease and the regional efforts used to test and report the disease cases. Wuhan, the center of the disease outbreak, and the entire Hubei province (of which Wuhan is the capital city), is considered different from other provinces in China in terms of the magnitude of the outbreak. In this study, we were able to obtain data from 1,493 confirmed cases outside Hubei province in China, including histories tracking these cases to the first reported contact with a disease carrier. Thus, the results of our study may help us to understand the behavior of the IP outside Hubei province. One potential factor that may contribute to the IP distribution is age. Figure A.1 of the Supplementary Material presents a plot of the reported mean (natural) logarithm of IP against mean age. Owing to such considerations, we consider the 27 provinces of China (all except Hubei) crossed by three age groups, namely, less than 23, 23–55, and greater than 55 years old, as small areas. This leads to a total of 64 small areas from which data are available.

The age factor is further considered as a covariate. Specifically, let y_i and x_i be the observed means of log(IP) and age, respectively, for the ith small area. We consider the logarithm of the IP because the lognormal distribution, along with other distributions of the exponential form in general, is widely used to characterize the distribution of the IP in investigations of infectious diseases; see, for example, Sartwell (1950). In other words, it is assumed that $\log(\mathrm{IP}) \sim N(\mu, \sigma^2)$. Therefore, it is quite natural to assume the dependence of log(IP) on the available auxiliary variable through its conditional mean functions. In addition, Figure A.1 suggests that the relationship between y_i and x_i may be quadratic. Furthermore, empirical evidence suggests that there is a potential climate effect. Specifically, provinces under continental climate conditions tend to have a lower IP than do those under non-continental climate conditions. Let I_i be an indicator, equal to one if the ith small area is associated with a continental-climate province, and zero otherwise. We consider the following area-level model:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 I_i + v_i + e_i, \quad i = 1, 2, \ldots, 64, \tag{5.1}$$

where v_i and e_i satisfy the conditions of the Fay−Herriot model. Here, D_i is

obtained via approximation to the sampling variance. That is, let σ^2 be the population variance of log(IP). Then, we have $D_i = \sigma^2/n_i \approx \hat{\sigma}^2/n_i$, where n_i is the sample size for the ith small area, of which y_i is the average log(IP), and $\hat{\sigma}^2$ is the sample variance of log(IP) obtained from the data combining all provinces and age groups. The D_i, along with other information, is provided in Table A.2 of the Supplementary Material.

The BPEs of the model parameters are $\hat{\beta}_0 = 2.518$, $\hat{\beta}_1 = -0.042$, $\hat{\beta}_2 = 0.001$, $\hat{\beta}_3 = -1.365$, and $\hat{A} = 0.062$. The estimate of β_2, the coefficient of x_i^2 in (5.1), is positive, whereas that of β_3, the coefficient of the continental-climate indicator, is negative. We then obtain the OBPs and their correspondng MSPE estimates using the MPR, PR, and JNR methods. Box plots of the MSPE estimates are presented in Figure A.2 of the Supplementary Material. It is seen that there are some negative values for the JNR estimates. Figure A.3 of the Supplementary Material presents the OBPs with corresponding margins of error, defined in the same way as in Figure 5. Note that some of the margins of error for JNR are not available owing to the negative MSPE estimates.

6. Conclusion

We have shown that the traditional PR method, originally derived under the correct specification of the underlying model, remains first-order unbiased under a misspecification of the mean function. Furthermore, we showed that the PR method can be (slightly) modified to achieve second-order unbiasedness. These results are important both theoretically and practically, because the PR method is popular in SAE and is easy to implement (available in commercial or online software packages).

Another way to understand the robustness of the PR method is to observe that, under the misspecified model, one can define the "true parameter" of β as β_* given by (2.8). Note that this is the parameter vector that minimizes the MSPE function; see Jiang, Nguyen and Rao (2011). The hypothetical "mean function" under β_* is then $x_i'\beta_*$, and the variation of the true small area mean, $\theta_i = x_i'\beta_0 + \Delta_i + v_i$, from the hypothetical mean is $\theta_i - x_i'\beta_* = x_i'(\beta_0 - \beta_*) + \Delta_i + v_i \equiv v_i^*$. The variance of v_i^* is exactly the right side of (2.9), owing to the independence of v_i and z_i (Assumption A1). Thus, the true variation is correctly captured by A_*, which is consistently estimated by the BPE.

Although the original PR method was derived under the assumption that the mean function is correctly specified, we showed that it remains first-order unbiased even if the mean function is misspecified. Note that the specific form

of the PR MSE estimator varies depending on the method used to estimate the model parameters. The original PR method (Prasad and Rao (1990)) was based on method of moments (MM) estimators of the model parameters; Datta and Lahiri (2000) derived the forms of the PR MSPE estimators with ML and REML estimators of the model parameters. We have derived the form of the PR MSPE estimator (without modification) using the BPE of the model parameters.

Furthermore, we has shown empirically that the PR method performs almost as well as a modified version that is second-order unbiased. In a way, this is similar to Lahiri and Rao (1995), who showed that although the PR MSPE estimator was derived under the normality assumption, it remains second-order unbiased when the random effects are not normal, provided that a certain moment condition holds. It is in this sense that we consider the PR method to be robust, thus justifying the title of the paper.

Supplementary Materials

The Supplementary Material contains proofs of the theoretical results, computational details of the simulation study, as well as additional tables and figure from the simulation study and real-data applications.

Acknowledgments

Xiaohui Liu's research was supported by NNSF of China (No. 11971208 and 11601197), and National Major Social Science Project of China (No. 21&ZD152). Haiqiang Ma's research was supported by NNSF of China (No. 12161042, 11961028 and 11701235), China Postdoctoral Science Foundation (No. 2019M662262), and NSF of Jiangxi Province (No. 20171BAB211004 and 20212BAB211010). The research of Jiming Jiang was supported by the NSF of the United States grant DMS-1510219. The authors are grateful to the two anonymous referees for their insightful comments and suggestions.

References

Battese, G. E., Harter, R. M. and Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.* **83**, 28–36.

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

Das, K., Jiang, J. and Rao, J. N. K. (2004). Mean squared error of empirical predictor. *Ann. Statist.* **32**, 818–840.

Datta, G. S., Kubokawa, T., Molina, I. and Rao, J. N. K. (2011). Estimation of mean squared

error of model-based small area estimators. *TEST* **20**, 367–388.

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. and Mandal, A. (2015). Small area estimation with uncertain random effects. *J. Amer. Statist. Assoc.* **110**, 1735–1744.

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika* **65**, 457–482.

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

Ganesh, N. (2009). Simultaneous credible intervals for small area estimation problems. *J. Multivariate Anal.* **100**, 1610–1621.

Guan, W. J., Ni, Z. Y., Hu, Y., Liang, W. H., Ou, C. Q., He, J. X. et al. (2020). Clinical characteristics of 2019 novel coronavirus infection in China. *New England J. Med.* **382**, 1708–1720.

Hall, P. and Maiti, T. (2006). On parametric bootstrap methods for small area prediction. *J. Roy. Statist. Soc. Ser. B (Statistical Methodology)* **68**, 221–238.

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

Jiang, J. (2010). *Large Sample Techniques for Statistics*. Springer, New York.

Jiang, J. (2017). *Asymptotic Analysis of Mixed Effects Models: Theory, Application and Open Problems*. Chapman & Hall/CRC.

Jiang, J., Lahiri, P. and Wan, S.-M. (2002). A unified jackknife theory for empirical best prediction with M-estimation. *Ann. Statist.* **30**, 1782–1810.

Jiang, J., Nguyen, T. and Rao, J. S. (2010). Fence method for nonparametric small area estimation. *Surv. Method.* **36**, 3–11.

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

Jiang, J., Nguyen, T. and Rao, J. S. (2015). Observed best prediction via nested-error regression with potentially misspecified mean and variance. *Surv. Method.* **41**, 37–55.

Jiang, J. and Rao, J. S. (2020). Robust small area estimation: An overview. *Annual Review of Statistics and Its Application* **7**, 337–360.

Jiang, J. and Torabi, M. (2020). Sumca: Simple, unified, Monte-Carlo assisted approach to second-order unbiased mean-squared prediction error estimation. *J. R. Statist. Soc. B (Statistical Methodology)* **82**, 467–485.

Jiongo, V. D., Haziza, D. and Duchesne, P. (2013). Controlling the bias of robust small area estimators. *Biometrika* **100**, 843–858.

Lahiri, P. and Rao, J. N. K. (1995). Robust estimation of mean squared error of small area estimators. *J. Amer. Statist. Assoc.* **90**, 758–766.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y. et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England J. Med.* **382**, 1199–1207.

Linton, N. M., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A. R., Jung, S. M. et al. (2020). Incubation period and other epidemiological characteristics of 2019 novel coronavirus Infections with right truncation: A statistical analysis of publicly available case data. *J. Clinical Med.* **9**, 538.

Morris, C. N. and Christiansen, C. L. (1995). Hierarchical models for ranking and for identifying extremes with applications. *Bayes Statistics*. Oxford University Press, London.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statist. Sci.* **28**, 40–68.

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of mean squared errors of small area estimators. *J. Amer. Statist. Assoc.* **85**, 163–171.

Rao, J. N. K. and Molina, I. (2015). *Small Area Estimation*. 2nd Edition. Wiley, New York.

Sartwell, P. E. (1950). The distribution of incubation periods of infectious diseases. *American Journal of Epidemiology* **51**, 310–318.

Sinha, S. K. and Rao, J. N. K. (2009). Robust small area estimation. *Can. J. Statist.* **37**, 381–399.

Song, P. X.-K. (2007). *Correlated Data Analysis: Modeling, Analytics and Applications*. Springer, New York.

Xiaohui Liu

Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330012, China.

E-mail: csuliuxh9122@gmail.com

Haiqiang Ma

Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330012, China.

E-mail: mahaiqiang@jxufe.edu.cn

Jiming Jiang

University of California, Davis, Davis, CA 95616, USA.

E-mail: jimjiang@ucdavis.edu

(Received August 2020; accepted March 2021)