Statistica Sinica

On Estimation of the Logarithm of the Mean Squared Prediction Error of A Mixed-effect Predictor

Jianling Wang[1], Thuan Nguyen[2], Yihui Luan[1], and Jiming Jiang[3]

Shandong University[1], *Oregon Health and Science* University[2], *and University of California,* Davis[3]

Abstract: The mean squared prediction error (MSPE) has been used as an important measure of uncertainty in small area estimation. It is desirable to produce a second-order unbiased MSPE estimator, that is, the bias of the estimator is $o(m^{-1})$, where m is the total number of small areas for which data are available. The task is difficult, however, especially if one needs to take into consideration that an MSPE estimator needs to be positive, or at least nonnegative. In fact, very few MSPE estimators have the property of being both second-order unbiased and guaranteed positive. We consider an alternative, easier approach of estimating the logarithm of the MSPE (log-MSPE), which avoids the issue of positivity. A second-order unbiased estimator of the log-MSPE is derived using the Prasad-Rao linearization method. Empirical studies demonstrate superiority of the proposed log-MSPE estimator over a naive log-MSPE estimator as well as an existing method known as McJack. A real-data example is considered.

Key words and phrases: bias-correction, log-MSPE, mixed effects, second-order unbiasedness, small area estimation, Taylor expansion

1. Introduction

The mean squared prediction error (MSPE) has been an important, and popular, measure of uncertainty in small area estimation (SAE; e.g., Rao and Molina 2015) ever since the seminal paper of Prasad and Rao (1990). It is desirable to produce a second-order unbiased estimator of the MSPE, that is, the bias of the MSPE estimator is $o(m^{-1})$, where m is the total number of small areas, for which data are available. See Liu, Ma and Jiang (2022a, b) for some recent advances. The mission is complicated, however, especially if one needs to take into account another desirable property of an MSPE estimator, that is, an MSPE estimator needs to be positive, or at least nonnegative. In fact, with very few exceptions (Prasad and Rao 1990, Chen and Lahiri 2011), all of the existing second-order unbiased MSPE estimators do not possess the double-property of being both second-order unbiased and guaranteed positive; see Jiang, Lahiri and Nguyen (2018, p. 408) for detailed discussion.

As noted by the latter authors, typically, it is fairly easy to obtain a positive MSPE estimator that is first-order unbiased. The complication arises when one tries to bias-correct the first-order unbiased MSPE estimator to make it second-order unbiased. This is because the resulting second-order unbiased MSPE estimator is no longer guaranteed positive. To deal with the latter drawback, one typically modify the value of the MSPE estimator when it is negative, for example, by truncating the estimator at zero, but by doing so it destroys the second-order unbiasedness. Jiang *et al.* (2018) used Hall and Maiti (2006) as an example to illustrate this dilemma. Intuitively, this may be compared an effort of trying to cover two ants, both moving fast in random directions, with two fingers of the same hand, which is difficult; however, the task is much easier with one finger covering just one ant, no matter how fast and randomly the latter moves.

Jiang *et al.* (2018) further proposed an alternative to "make life easier" by estimating the logarithm of the MSPE (log-MSPE), instead of the MSPE itself. They noted a number of advantages of targeting the log-MSPE, including the latter being a simple, one-to-one transformation of the MSPE that one can easily convert a log-MSPE estimator to an MSPE estimator by taking the exponential, that reporting the log-MSPE results is often space-saving, some advantage in hypothesis testing, a linear association between the logarithms of MSPE and the square root of MSPE and, more importantly, that one does not need to worry about the issue of positivity, because the exponential of a log-MSPE estimator is always positive. Jiang *et al.* (2018) further proposed a Monte-Carlo jackknife method of estimating the log-MSPE, called McJack. The authors showed that McJack produces a second-order unbiased estimator of the log-MSPE.

In the SAE literature, there are two standard methods of producing a second-order unbiased MSPE estimator, namely, the Prasad-Rao linearization method (Prasad and Rao 1990) and resampling method (e.g., Jiang *et al.* 2002, Hall and Maiti 2006); also see Rao and Molica (2015). The McJack belongs to the resampling methods. The main objective of this paper is to develop a class of second-order unbiased estimators of the log-MSPE using the linearization method, and demonstrate its advantage over the existing methods.

The method is described, in general, in Section 2. In Section 3 we consider a special case of estimating the log-MSPE of the empirical best predictor (EBP) based on generalized linear mixed models (GLMM; e.g., Jiang and Nguyen 2021). Some simulation results are presented in Section 4. A real-data example is discussed in Section 5. A summary and concluding remark are offered in Section 6.

2. Second-order unbiased log-MSPE estimator

Let θ denote a mixed effect of interest, which may be a small area mean, and $\hat{\theta}$ a predictor of θ. For example, $\hat{\theta}$ may be the empirical best linear unbiased predictor (EBLUP; e.g., Rao and Molina 2015), or observed best predictor (OBP; Jiang, Nguyen and Rao 2011). Define the MSPE of $\hat{\theta}$ as

$$\mathrm{MSPE} \equiv \mathrm{MSPE}(\hat{\theta}) = \mathrm{E}(\hat{\theta} - \theta)^2. \tag{2.1}$$

Let $\widetilde{\mathrm{MSPE}}$ be an estimator of the MSPE that possesses the below properties:

(i) $\widetilde{\mathrm{MSPE}}$ is positive with probability one;

(ii) $\widetilde{\mathrm{MSPE}}$ is at least first-order unbiased; that is, $\mathrm{E}(\widetilde{\mathrm{MSPE}} - \mathrm{MSPE}) = O(m^{-1})$; and

(iii) $\widetilde{\mathrm{MSPE}} - \mathrm{MSPE} = O_{\mathrm{P}}(m^{-1/2})$

(see, e.g., Jiang 2010, sec. 3.4 for the definition of O_{P} and also o_{P}). More specifically, suppose that we have the following expressions:

$$\mathrm{MSPE} = a(\psi) + o(1), \tag{2.2}$$

$$\mathrm{E}(\widetilde{\mathrm{MSPE}} - \mathrm{MSPE}) = m^{-1}b(\psi) + o(m^{-1}), \tag{2.3}$$

$$\mathrm{E}(\widetilde{\mathrm{MSPE}} - \mathrm{MSPE})^2 = m^{-1}c(\psi) + o(m^{-1}), \tag{2.4}$$

where $a(\cdot), b(\cdot), c(\cdot)$ are continuous, which may depend on m, but $a(\psi)$, $b(\psi)$, $c(\psi)$ are bounded and $a(\psi)$ has a positive lower bound for the ψ

that satisfies (2.2)–(2.4). Note that (2.4) is a consequences of (iii) under regularity conditions, By Taylor series expansion, we have

$$\log(\widetilde{\mathrm{MSPE}}) - \log(\mathrm{MSPE}) = \frac{\widetilde{\mathrm{MSPE}} - \mathrm{MSPE}}{\mathrm{MSPE}} - \frac{(\widetilde{\mathrm{MSPE}} - \mathrm{MSPE})^2}{2\mathrm{MSPE}^2} + O_{\mathrm{P}}(m^{-3/2}).$$

Thus, under regularity conditions, it can be shown that

$$\mathrm{E}\{\log(\widetilde{\mathrm{MSPE}}) - \log(\mathrm{MSPE})\} = \frac{2a(\psi)b(\psi) - c(\psi)}{2ma^2(\psi)} + o(m^{-1}). \tag{2.5}$$

Let $\hat{\psi}$ be a consistent estimator of ψ. Then, under regularity conditions, we have

$$\begin{aligned} \frac{2a(\psi)b(\psi) - c(\psi)}{2ma^2(\psi)} &= \mathrm{E}\left\{\frac{2a(\hat{\psi})b(\hat{\psi}) - c(\hat{\psi})}{2ma^2(\hat{\psi})}\right\} \\ &\quad - \frac{1}{m}\mathrm{E}\left\{\frac{2a(\hat{\psi})b(\hat{\psi}) - c(\hat{\psi})}{2a^2(\hat{\psi})} - \frac{2a(\psi)b(\psi) - c(\psi)}{2a^2(\psi)}\right\} \\ &= \mathrm{E}\left\{\frac{2a(\hat{\psi})\bar{b}(\hat{\psi}) - \bar{c}(\hat{\psi})}{2a^2(\hat{\psi})}\right\} + o(m^{-1}), \end{aligned} \tag{2.6}$$

where $\bar{b}(\psi) = b(\psi)/m$ and $\bar{c}(\psi) = c(\psi)/m$. Note that $\{2a(\psi)b(\psi) - c(\psi)\}/2ma^2(\psi)$ is non-random. Combining (2.5) and (2.6), we have

$$\mathrm{E}\{\log(\widetilde{\mathrm{MSPE}}) - \log(\mathrm{MSPE})\} = \mathrm{E}\left\{\frac{2a(\hat{\psi})\bar{b}(\hat{\psi}) - \bar{c}(\hat{\psi})}{2a^2(\hat{\psi})}\right\} + o(m^{-1}). \tag{2.7}$$

Thus, if we define a bias-corrected log-MSPE estimator as

$$\widehat{\log}(\mathrm{MSPE}) = \log(\widetilde{\mathrm{MSPE}}) - \frac{2a(\hat{\psi})\bar{b}(\hat{\psi}) - \bar{c}(\hat{\psi})}{2a^2(\hat{\psi})}, \tag{2.8}$$

then, by (2.7), we have

$$\mathrm{E}\{\widehat{\log}(\mathrm{MSPE}) - \log(\mathrm{MSPE})\} = o(m^{-1}); \tag{2.9}$$

thus, $\widehat{\log}(\mathrm{MSPE})$ is a second-order unbiased estimator of log(MSPE).

The derivation above is quite general, which, depending on the specifications of $\widetilde{\mathrm{MSPE}}$, and $a(\cdot), \bar{b}(\cdot), \bar{c}(\cdot)$, leads to a class of second-order unbiased estimators of the log-MSPE. Note that, however, the exponential of a second-order unbiased log-MSPE estimator is not necessarily a second-order unbiased MSPE estimator. This is because the back-transformation (exponential) results in a bias of $O(m^{-1})$.

Next, we demonstrate the method by considering a special case.

3. EBP based on GLMM

In the context of SAE with discrete or categorical responses, Jiang (2003) proposed an EBP method based on a GLMM, which assumes that, conditional on random effect vectors, $v_i = (v_{ij})_{1 \le j \le n_i}, 1 \le i \le m$, responses $y_{ij}, 1 \le j \le n_i$ are independent with conditional pmf, or pdf, in the form of

$$f(y_{ij}|v_i) = \exp\left\{\left(\frac{w_{ij}}{\phi}\right)(y_{ij}\xi_{ij} - r(\xi_{ij})) + s\left(y_{ij}, \frac{\phi}{w_{ij}}\right)\right\},$$

where $r(\cdot), s(\cdot,\cdot)$ are functions associated with the exponential family (McCullagh and Nelder 1989, ch. 2); ϕ is a dispersion parameter, which in some cases is known; w_{ij} is a weight such that $w_{ij} = 1$ for ungrouped data, $w_{ij} = l_{ij}$ for grouped data if the average is considered as response (l_{ij} is the group size), and $w_{ij} = l_{ij}^{-1}$ if the sum of individual responses is considered. Furthermore, ξ_{ij} is associated with a linear predictor, $\eta_{ij} = x'_{ij}\beta + z'_{ij}v_i$ through a link function, $g(\xi_{ij}) = \eta_{ij}$, or $\xi_{ij} = h(\eta_{ij})$, where $h = g^{-1}$. Here, $x_{ij} = (x_{ijk})_{1\le k\le p}$ and $z_{ij} = (z_{ijk})_{1\le k\le r}$ are known vectors, and β is a vector of regression coefficients. In case of a canonical link, one has $\xi_{ij} = \eta_{ij}$. Finally, $v_1, \dots, v_m$ are independent with density $f_\nu(\cdot)$, where ν is a vector of variance components. For simplicity, we focus on cases where ϕ is known. This includes important cases such as the binomial and Poisson families. Let $\psi = (\beta', \nu')'$.

Consider prediction a possibly nonlinear mixed effect in the form of

$$\zeta = \zeta(\beta, v_S)\ , \tag{3.10}$$

where S is a subset of $\{1, \dots, m\}$, and $v_S = (v_i)_{i\in S}$. Let $y_S = (y_i)_{i\in S}$, where $y_i = (y_{ij})_{1\le j\le n_i}$; and $y_{S-} = (y_i)_{i\notin S}$. According to Jiang (2003), the best predictor (BP) of ζ in the sense of minimum MSPE, is given by

$$\begin{aligned}\tilde{\zeta} &= \frac{\int \zeta(\beta, v_S)\exp(\phi^{-1}\sum_{i\in S} s_i(\beta, v_i))\prod_{i\in S} f_\nu(v_i)\prod_{i\in S} dv_i}{\prod_{i\in S}\int \exp(\phi^{-1}s_i(\beta, v))f_\nu(v)dv} \\ &\equiv u(y_S, \psi), \end{aligned} \tag{3.11}$$

where $s_i(\beta, v) = \sum_{j=1}^{n_i} w_{ij}[y_{ij}h(x'_{ij}\beta + z'_{ij}v) - r(h(x'_{ij}\beta + z'_{ij}v))]$. The integral involved in (3.11) may be evaluated via numerical integration or Monte-Carlo methods. As for the unknown parameters, ψ, involved in (3.11), Jiang (2003) suggests to use the method of moments (MoM) estimators, which are consistent (Jiang 1998). Let $\hat{\psi}$ denote the MoM estimator of ψ. If one replaces the ψ in (3.11) by $\hat{\psi}$, one gets the empirical BP, or EBP,

$$\hat{\zeta} \quad = \quad u(y_S, \hat{\psi}). \tag{3.12}$$

The MSPE of the EBP is of primary concern. Jiang (2003) derived a second-order unbiased MSPE estimator, which is not guaranteed positive. We now apply the general result of Section 2 to derive a second-order unbiased log-MSPE estimator. Suppose that

$$\mathrm{E}(\hat{\psi} - \psi)(\hat{\psi} - \psi)' = m^{-1}V(\psi) + o(m^{-1}). \tag{3.13}$$

Then, according to Jiang (2003), one has the following expression:

$$\mathrm{MSPE} \equiv \mathrm{MSPE}(\hat{\zeta}) = d(\psi) + m^{-1}e(\psi) + o(m^{-1}), \tag{3.14}$$

where $d(\psi) = \mathrm{MSPE}(\tilde{\zeta}) = \mathrm{E}(\zeta^2) - \mathrm{E}(\tilde{\zeta}^2)$, using the fact that $\tilde{\zeta} = \mathrm{E}(\zeta|y)$, and $e(\psi) = \mathrm{E}\{(\partial u/\partial \psi')V(\psi)(\partial u/\partial \psi)\}$. We now obtain a further expression for $V(\psi)$. According to Jiang (1998), $\hat{\psi}$ is a solution to the estimating equation

$$M(\psi) = \hat{M}, \tag{3.15}$$

where $\hat{M} = (\hat{M}_k)_{1 \le k \le q}$, with $q = \dim(\psi)$, is a vector of normalized statistics in the sense that, when ψ is the true parameter vector, one has $\mathrm{E}(\hat{M}) = O(1)$ and $\mathrm{Var}(\hat{M}) = O(m^{-1})$; $M(\psi) = [M_k(\psi)]_{1 \le k \le q}$ with $M_k(\psi) = \mathrm{E}_\psi(\hat{M}_k)$. It is known that, under regularity conditions, $\hat{\psi}$ is root-m consistent (Jiang 1998), that is, $\hat{\psi} - \psi = O_{\mathrm{P}}(m^{-1/2})$. Write $M = M(\psi)$ when ψ is the true parameter vector. Then, by Taylor series expansion at ψ, the true parameter vector, one has, under regularity conditions,

$$\hat{M} = M(\hat{\psi}) = M + A(\hat{\psi} - \psi) + O_{\mathrm{P}}(m^{-1}), \tag{3.16}$$

where $A = \partial M / \partial \psi'$. (3.16) implies the following asymptotic expansion

$$\hat{\psi} - \psi = A^{-1}(\hat{M} - M) + O_{\mathrm{P}}(m^{-1}), \tag{3.17}$$

which results in, under regularity conditions, the following approximation:

$$\mathrm{E}(\hat{\psi} - \psi)(\hat{\psi} - \psi)' = A^{-1}\mathrm{E}(\hat{M} - M)(\hat{M} - M)'(A^{-1})' + o(m^{-1}). \tag{3.18}$$

Note that $\mathrm{E}(\hat{M} - M)(\hat{M} - M)' = \mathrm{Var}(\hat{M})$. This leads to a further expression:

$$V(\psi) = mA^{-1}\mathrm{Var}(\hat{M})(A^{-1})'. \tag{3.19}$$

Thus, combining (3.14) and (3.19), we have

$$\mathrm{MSPE} = d(\psi) + b_3(\psi) + o(m^{-1}) = d(\psi) + o(1), \tag{3.20}$$

where $b_3(\psi) = \mathrm{E}\{(\partial u/\partial\psi')A^{-1}\mathrm{Var}(\hat{M})(A^{-1})'(\partial u/\partial\psi)\}$. In fact, $b_3(\psi) = O(m^{-1})$. It follows that (2.2) holds with $a(\psi) = d(\psi)$.

Now define $\widetilde{\mathrm{MSPE}} = d(\hat{\psi})$. By the definition of $d(\psi)$, condition (i) of Section 2 is satisfied (assuming non-singularity). Also, by (3.20), we have

$$\mathrm{E}(\widetilde{\mathrm{MSPE}} - \mathrm{MSPE}) = \mathrm{E}\{d(\hat{\psi}) - d(\psi)\} - b_3(\psi) + o(m^{-1}). \quad (3.21)$$

Furthermore, again by Taylor series expansion at the true ψ, and (3.17), it can be shown that

$$\begin{aligned} d(\hat{\psi}) - d(\psi) &= \frac{\partial d}{\partial\psi'}(\hat{\psi} - \psi) + \frac{1}{2}(\hat{M} - M)'(A^{-1})'\frac{\partial^2 d}{\partial\psi\partial\psi'}A^{-1}(\hat{M} - M) \\ &\quad + o_{\mathrm{P}}(m^{-1}). \end{aligned} \quad (3.22)$$

We can expand (3.17) to obtain a further expansion (see the supplement):

$$\begin{aligned} \hat{\psi} - \psi &= A^{-1}(\hat{M} - M) \\ &\quad - \frac{1}{2}A^{-1}\left[(\hat{M} - M)'(A^{-1})'B_k A^{-1}(\hat{M} - M)\right]_{1\le k\le q} \\ &\quad + o_{\mathrm{P}}(m^{-1}), \end{aligned} \quad (3.23)$$

where $B_k = \partial^2 M_k/\partial\psi\partial\psi'$. Combining (3.22) and (3.23), we have, under regularity conditions, that

$$\mathrm{E}\{d(\hat{\psi}) - d(\psi)\} = \frac{b_1(\psi) - b_2(\psi)}{2} + o(m^{-1}), \quad (3.24)$$

where $b_1(\psi) = \mathrm{E}\{(\hat{M} - M)'(A^{-1})'(\partial^2 d/\partial\psi\partial\psi')A^{-1}(\hat{M} - M)\}$ and $b_2(\psi) = (\partial d/\partial\psi')A^{-1}[\mathrm{E}(\hat{M} - M)'(A^{-1})'B_k A^{-1}(\hat{M} - M)]_{1\le k\le q}$.

Combining (3.21), (3.24), it follows that (2.3) holds with

$$\bar{b}(\psi) = \frac{b_1(\psi) - b_2(\psi)}{2} - b_3(\psi).$$

Finally, by (3.20), (3.22), (3.23), it can be shown that

$$\widetilde{\text{MSPE}} - \text{MSPE} = \frac{\partial d}{\partial \psi'} A^{-1}(\hat{M} - M) + O_{\text{P}}(m^{-1}). \tag{3.25}$$

(3.25) implies that, under regularity conditions, (2.4) holds with

$$\bar{c}(\psi) = \frac{\partial d}{\partial \psi'} A^{-1} \text{Var}(\hat{M})(A^{-1})' \frac{\partial d}{\partial \psi}.$$

In conclusion, the general result of Section 2 applies with $\widetilde{\text{MSPE}} = d(\hat{\psi})$, $a(\psi) = d(\psi)$, $\bar{b}(\psi)$ and $\bar{c}(\psi)$ specified above and below (3.25), respectively.

The following expressions are computationally more convenient:

$$b_1(\psi) = \text{tr}\left((A^{-1})' \frac{\partial^2 d}{\partial \psi \partial \psi'} A^{-1} \text{Var}(\hat{M})\right), \tag{3.26}$$

$$b_2(\psi) = \frac{\partial d}{\partial \psi'} A^{-1} \left[\text{tr}\left((A^{-1})' B_k A^{-1} \text{Var}(\hat{M})\right)\right]_{1 \le k \le q}, \tag{3.27}$$

$$b_3(\psi) = \text{tr}\left(A^{-1} \text{Var}(\hat{M})(A^{-1})' \text{E}\left(\frac{\partial u}{\partial \psi} \frac{\partial u}{\partial \psi'}\right)\right), \tag{3.28}$$

where the expectation inside the trace is with respect to the y_S in (3.11).

Computational/practical notes:

1. Although, in theory, $d(\psi) = \text{MSPE}(\tilde{\zeta})$ should be positive for any ψ, depending on the method used to evaluate it, the value of $d(\psi)$ can occasionally be negative. For example, in the next section we use numerical integration to evaluate $d(\psi)$. Then, due to the integral approximations, $d(\hat{\psi})$ can occasionally take negative values. When the value of $d(\hat{\psi})$ is negative, we suggest to evaluate it via a Monte-Carlo method as in Jiang *et al.* (2018). The latter is computationally more time-consuming than numerical integration, but is guaranteed to produce a positive number.

2. Also, the matrix A can occasionally be singular. In this case, we suggest to use the Moore-Penrose generalized inverse of A in place of A^{-1}.

3. In some cases there are known bounds for the value of MSPE. Such bounds should be used, in practice, to improve the precision of the log-MSPE estimate. For example, in the case considered in the next section, the MSPE is bounded by 1, hence the log-MSPE is bounded by 0. Thus, the value of the log-MSPE estimate is taken as 0 (hence the MSPE estimate equal to 1) in case it is greater than 0.

4. Example and simulation

Consider a case of mixed logistic model for small area estimation (e.g., Jiang and Lahiri 2001). Suppose that conditional on p_i, y_{ij}'s are independent *Bernoulli* with $P(y_{ij} = 1|p_i) = p_i$, $i = 1, \cdots, m$, $j = 1, \cdots, k_i$. Also, we have $\text{logit}(p_i) = \log(p_i/(1-p_i)) = \mu + v_i$, where μ is an known parameter. Furthermore, $v_1, \ldots, v_m$ are independent random effects. Two distributions of the random effects are considered: (a) $v_i \sim N(0, \sigma^2)$, where σ^2 is an unknown variance; and (b) $v_i \sim \mathrm{LP}(\sigma)$, where $\mathrm{LP}(\sigma)$ denotes the Laplace distribution with pdf $f(x|\sigma) = (2\sigma)^{-1} e^{-|x|/\sigma}$, $-\infty < x < \infty$.

For simplicity, let $k_i = k > 1, 1 \le i \le m$. It is convenient to use the expression $v_i = \sigma\xi_i$, where $\xi_i \sim N(0,1)$ in case (a), and $\xi_i \sim \mathrm{LP}(1)$ in case (b). Consider prediction of the conditional probability, $p_i = h(\mu + \sigma\xi_i)$, where $h(x) = e^x/(1+e^x)$. According to Jiang (2003), the BP of p_i is

$$\tilde{p}_i = e^{\mu} \frac{\mathrm{E}\{\exp((y_{i\cdot} + 1)\sigma\xi - (k+1)\log(1 + e^{\mu+\sigma\xi}))\}}{\mathrm{E}\{\exp(y_{i\cdot}\sigma\xi - k\log(1 + e^{\mu+\sigma\xi}))\}} \equiv u(y_{i\cdot}, \psi), \quad (4.29)$$

where $\psi = (\mu, \sigma)'$, $y_{i\cdot} = \sum_{j=1}^{k} y_{ij}$, and the expectations are with respect to ξ, which is $N(0,1)$ in case (a) and $\mathrm{LP}(1)$ in case (b). The EBP, $\hat{p}_i$, is $\tilde{p}_i$ with ψ replaced by $\hat{\psi}$, the MoM estimator. The latter is the solution to (3.15) with $q = 2$, $\hat{M}_1 = (mk)^{-1} y_{\cdot\cdot}$, where $y_{\cdot\cdot} = \sum_{i=1}^{m}\sum_{j=1}^{k} y_{ij}$, $\hat{M}_2 = \{mk(k-1)\}^{-1} \sum_{i=1}^{m} (y_{i\cdot}^2 - y_{i\cdot})$, and $M_s(\psi) = \mathrm{E}\{h^s(\mu + \sigma\xi)\}, s = 1, 2$ (Jiang 1998). We have the following expression (see the supplementary material):

$$\begin{aligned} d(\psi) &= \mathrm{E}\{h^2(\mu + \sigma\xi)\} \\ &- \sum_{l=0}^{k} u^2(l, \psi) \binom{k}{l} \mathrm{E}\left\{\exp\left(l(\mu + \sigma\xi) - k\log\left(1 + e^{\mu+\sigma\xi}\right)\right)\right\}, \end{aligned} \quad (4.30)$$

where $u(l, \psi)$ is $u(y_{i\cdot}, \psi)$ [see (4.29)] with $y_{i\cdot} = l$.

For notation simplicity, write $h = h(\mu + \sigma\xi)$ when ψ is the true parameter vector and ξ is as above. Similarly, we write $h' = h'(\mu + \sigma\xi)$,

$h'' = h''(\mu + \sigma\xi)$, and $g = (h')^2 + hh''$. It is easy to derive the following:

$$A = \begin{bmatrix} \mathrm{E}(h') & \mathrm{E}(h'\xi) \\ 2\mathrm{E}(hh') & 2\mathrm{E}(hh'\xi) \end{bmatrix}, \quad B_1 = \begin{bmatrix} \mathrm{E}(h'') & \mathrm{E}(h''\xi) \\ \mathrm{E}(h''\xi) & \mathrm{E}(h''\xi^2) \end{bmatrix},$$

$$B_2 = 2 \begin{bmatrix} \mathrm{E}(g) & \mathrm{E}(g\xi) \\ \mathrm{E}(g\xi) & \mathrm{E}(g\xi^2) \end{bmatrix}.$$

Expressions of the elements of $\mathrm{Var}(\hat{M})$ are given in Section A.2.2 of the supplement; those of the partial derivatives involved in (3.26)–(3.28) are given in Section A.2.3 of the supplement. Note that, in this case, we have

$$\mathrm{E}\left(\frac{\partial u}{\partial \psi}\frac{\partial u}{\partial \psi'}\right) = \sum_{l=0}^{k} \binom{k}{l} \left.\frac{\partial u}{\partial \psi}\frac{\partial u}{\partial \psi'}\right|_{(l,\psi)} \mathrm{E}\{s(l, k, \mu + \sigma\xi)\},$$

where $s(a, b, w) = \exp(aw - b\log(1 + e^w))$. All of the expectations involved were evaluated via numerical integration using the integrate() function in R (lower bound $= -5$; upper bound $= 5$). Also note that, in this case, the MSPE is naturally bounded by 1, hence the log-MSPE bounded by 0. Thus, the value of the log-MSPE estimate is taken as 0 in case it is positive (see Note 3 at the end of the last section).

Simulation studies are carried out to evaluate performance of the bias-corrected log-MSPE estimator, (2.8). The latter is compared with a naive log-MSPE estimator, which is simply $\log(\widetilde{\mathrm{MSPE}})$, the first term on the right side of (2.8). Consider prediction of p_1 via the EBP. We consider

$m = 25, 50, 100$ and $k_i = 4, 1 \leq i \leq m$. The true parameters are $\mu = -1.0$ and $\sigma = 2.0$. The Monte-Carlo sample size used to evaluate $d(\hat{\psi})$, when it is occasionally negative (see Note 1 at the end Section 3), is $N_{\text{mc}} = 1,000$.

There performance measures are considered:

(1) Bias, E(log-MSPE estimator)$-$log-MSPE;

(2) Percentage relative bias (%RB), which is $100 \times$ (Bias/|log-MSPE|); and

(3) Coefficient of variation (CV), which is the standard deviation (s.d.) of the log-MSPE estimator divided by the absolute value of the mean of the log-MSPE estimator.

Here, the mean and s.d. are the simulated mean and s.d., respectively, and the (true) MSPE is evaluated via the simulation runs.

Results, based on $N_{\text{sim}} = 2,000$ simulation runs, are presented in Table 1. The bias-corrected estimator, $\widehat{\log}(\text{MSPE})$, appears to be a clear winner, especially in terms of %RB, in both case (a) and case (b).

Next, we compare our log-MSPE estimator with the McJack estimator of Jiang *et al.* (2018). The latter is also intended to estimate the log-MSPE. Because McJack is computationally intensive, and the computational burden increases quickly as m increases, the comparison is limited to the case of $m = 25, k = 4$. In addition to the above performance measures, we also consider the average computing time (ACT; in seconds) per simulation run. As

Table 1: **Comparison with Naive log-MSPE estimator**

Case	Sample Size	Simulated log-MSPE	$\log(\widetilde{\mathrm{MSPE}})$ Bias	%RB	CV	$\widehat{\log}(\mathrm{MSPE})$ Bias	%RB	CV
(a)	$m = 25, k = 4$	-3.54	-0.34	-9.64	0.50	-0.03	-0.87	0.08
	$m = 50, k = 4$	-3.63	-0.10	-2.70	0.02	-0.02	-0.46	0.02
	$m = 100, k = 4$	-3.68	-0.03	-0.90	0.02	0.00	0.12	0.02
(b)	$m = 25, k = 4$	-3.61	-0.22	-6.09	0.19	-0.02	-0.46	0.07
	$m = 50, k = 4$	-3.66	-0.10	-2.76	0.02	-0.03	-0.86	0.03
	$m = 100, k = 4$	-3.71	-0.03	-0.74	0.02	0.01	0.14	0.02

McJack depends on the Monte-Carlo (MC) sample size used in evaluating expectations, we consider two difference MC sample sizes, $K_{\mathrm{mc}} = 50, 100$. Due to the computational intensity of McJack, here we set $N_{\mathrm{sim}} = 500$ (instead of $N_{\mathrm{sim}} = 2,000$, as in the previous case). The results are reported in Table 2, in which the results for $\widehat{\log}(\mathrm{MSPE})$, with the exception of ACT, are copied from Table 1.

It is seen that $\widehat{\log}(\mathrm{MSPE})$ does better in terms of both %RB and CV, although the results are comparable. The biggest difference is in terms of computational efficiency, in which $\widehat{\log}(\mathrm{MSPE})$ is doing much better. For

Table 2: **Comparison with McJack**

Case	Method	Bias	%RB	CV	ACT
(a)	$\widehat{\log}$(MSPE)	-0.03	-0.87	0.08	0.03
	McJack ($K_{\text{mc}} = 50$)	-0.07	-1.97	0.14	13.14
	McJack ($K_{\text{mc}} = 100$)	-0.05	-1.32	0.11	25.71
(b)	$\widehat{\log}$(MSPE)	-0.02	-0.46	0.07	0.02
	McJack ($K_{\text{mc}} = 50$)	-0.04	-1.09	0.14	9.69
	McJack ($K_{\text{mc}} = 100$)	-0.03	-0.73	0.11	19.57

example, in case (a), the ACT of McJack is 438 times that of $\widehat{\log}$(MSPE) when $K_{\text{mc}} = 50$, and 857 times that of $\widehat{\log}$(MSPE) when $K_{\text{mc}} = 100$. Keep in mind that, due to the computational intensity, here we only consider the case of $m = 25$. When m is larger, the computing time needed for McJack may become unbearable, especially if one needs results quickly. This leaves $\widehat{\log}$(MSPE) as the only feasible method that is capable in producing a second-order unbiased log-MSPE estimator when m is large.

5. A real-data example

Brooks et al. (1997) presented six datasets recording fetal mortality in mouse litters. As an application, we consider the HS2 dataset from Table 4 of their paper, which reports the number of dead implants in litters of mice from untreated experimental animals. Jiang and Zhang (2001) analyzed the data based on a GLMM (also see Jiang and Nguyen 2021, sec. 4.4.1). Let $y_{ij},\ i = 1, ..., m,\ j = 1, \dots, k_i$ be binary responses such that $y_{ij} = 1$ if the jth implant in the ith litter is dead, and $y_{ij} = 0$ otherwise. Here, $m = 1,328$ is the total number of litters. The y_{ij}s are assumed to satisfy the same mixed logistic model with normally distributed random effects, as described at the beginning of Section 4. We have also considered the mixed logistic model with Laplacian random effects, again as described in Section 4. The results are very similar and therefore omitted.

Note that the data are unbalanced in this case (i.e., the k_is are not equal). Thus, the definition of $\hat{M}_s, s = 1, 2$ are different than those in the previous section. Specifically, we have $\hat{M}_1 = k_\cdot^{-1} y_{\cdot\cdot}$, where $k_\cdot = \sum_{i=1}^m k_i$, and $\hat{M}_2 = \{\sum_{i=1}^m k_i(k_i - 1)\}^{-1} \sum_{i=1}^m (y_{i\cdot}^2 - y_{i\cdot})$. According to Jiang and Nguyen (2021, sec. 4.4.1), the MoM estimates are $\hat{\mu} = -2.276$ and $\hat{\sigma} = 0.644$. Expression of $\mathrm{Var}(\hat{M})$ in this case, as well as additional expressions in terms of the current data structure, are given in Section A.3 of the supplement.

Once again, we are interested in predicting the conditional probability, $p_i = h(\mu + \sigma\xi_i)$, for all $m = 1,328$ litters. The values of the EBP, as well as the corresponding log-MSPE estimates, only depend the values of $\hat{\mu}, \hat{\sigma}, k_i$

and $y_{i\cdot}$. In this case, the McJack estimates are very computational intensity to compute ($m = 1,328$ in this case!); see discussion in the last paragraph of Section 4. On the other hand, it is fairly easy to obtain the log-MSPE estimates using our method. As in the previous section, the EBP and log-MSPE estimates are computed via numerical integration. The results, including the EBP and corresponding square root of the MSPE (RMSPE) estimate, obtained via simple transformation from the log-MSPE estimate, are reported in Table 3. The table is constructed in a way similar to Table 4 of Brooks et al. (1997). The RMSPE is often used as a measure of uncertainty in a way similar to the standard error in parameter estimation.

It is seen that the EBP decreases as k_i increases, and increases as $y_{i\cdot}$ increases. While both trends can be shown theoretically, there are also intuitive explanations. Recall the EBP predicts the conditional probability that the implant is dead given the observed count, $y_{i\cdot}$. For example, take a look at $k_i = 7$. If $y_{i\cdot}$ is 0, that is, no implant is dead, the predicted probability death is 0.084. If $y_{i\cdot} = 1$, that is, one implant is dead, one would expect the conditional probability of death to increase; this is indeed the case as the predicted probability of death is now 0.112. Now let $y_{i\cdot}$ be fixed, say, $y_{i\cdot} = 1$. As k_i increases, one expects more death; therefore, the probability of exactly one death should decrease.

Table 3: **Analysis of Mice Mortality Data:** For Each # of Implants, 1st Row Is Observed # of Cases; 2nd Row Is RMSPE (Column RMSPE) and EBPs (Columns 0–9)

# of implants (k_i)	# of dead implants ($y_{i\cdot}$)										
	RMSPE	0	1	2	3	4	5	6	7	8	9
1		15	1								
	0.062	0.103	0.144								
2		6	1	2							
	0.061	0.099	0.137	0.184							
3		6	6								
	0.060	0.096	0.131								
4		7	2	3		2					
	0.059	0.092	0.125	0.166		0.27					
5		16	9	3	3	1					
	0.057	0.089	0.121	0.159	0.203	0.255					
6		57	38	17	2	2					
	0.056	0.087	0.116	0.152	0.194	0.241					
7		119	81	45	6	1			1		
	0.056	0.084	0.112	0.146	0.185	0.23			0.385		
8		173	118	57	16	3				1	
	0.055	0.082	0.109	0.141	0.178	0.219				0.417	
9		136	103	50	13	6	1	1			
	0.054	0.08	0.106	0.136	0.171	0.210	0.252	0.298			
10		54	51	32	5	1					1
	0.053	0.078	0.102	0.131	0.164	0.201					0.425
11		13	15	12	3	1					
	0.052	0.076	0.100	0.127	0.159	0.194					
12			4	3	1						
	0.051		0.097	0.123	0.153						
13				1					1		
	0.051			0.120					0.290		

As for RMSPE, first note that it only depends on k_i. This is reasonable because the MSPE is unconditional, that is, it does not depend on the value of $y_{i\cdot}$. In fact, under the assumed model, $y_{i\cdot}, i = 1, \dots, m$ are i.i.d., whose distribution only depends on k_i and ψ. It is also observed that the RMSPE decreases as k_i increases. This also makes sense because k_i is part of the sample sizes. As k_i increases, more information is available for better prediction; as a result, the MSPE should decrease.

6. Discussion and concluding remarks

We have derived a linearization-based method for producing a second-order unbiased estimator of the log-MSPE of a predictor of a mixed effect of interest. We apply the method to the special case of predicting a (possibly) non-linear mixed effect via EBP under a GLMM. We demonstrate the superiority of our method over a naive predictor and the McJack, especially in terms of the computational efficiency over the latter. We use a real-data example to illustrate the practical relevance of our method.

The computational disadvantage of McJack makes it difficult to evaluate its performance via large-scale simulation studies, in which a large number of simulation runs need to be carried out in order to produce accurate results, even when m is moderately large. it may also be inconvenient in practice when measure-of-uncertainty results need to be produced in a timely manner. Our proposed log-MSPE estimator does not have any of these issues, and it is as accurate as McJack, if not more accurate.

Having said that, the current approach is similar to the Prasad-Rao linearization method in estimating the MSPE (Prasad and Rao 1990); thus, it does not result in any simplification in terms of the analytic derivations, compared to the Prasad-Rao method. There is, however, a potentially middle ground between the analytically tedious Prasad-Rao method and the computationally intensive McJack method. Recently, Jiang and Torabi (2020) proposed a Sumca method for estimating the MSPE of a complex predictor. The method may be viewed as a hybrid of the linearization and resampling methods. Namely, it uses the linearization method to obtain a leading term of the MSPE estimator, and a Monte-Carlo method to obtain a bias-correction to achieve the second-order unbiasedness. The linearization is (much) simpler to derive compared to the Prasad-Rao (because one does not need to achieve the second-order unbiasedness for the leading term), and the Monte-Carlo bias-correction is computationally much faster than McJack or double bootstrap (Hall & Maiti 2006). In our future work we shall explore extending the Sumca method to the log-MSPE estimation.

As mentioned in Section1, most existing second-order unbiased MSPE estimators may take negative values. In particular, Liu *et al.* (2022b) proposed a modified Prasad-Rao (PR) estimator for estimating the MSPE of the OBP (Jiang *et al.* 2011). Empirical studies suggest that the modified PR estimator does not take negative values; however, so far there is no proof showing that this estimator is guaranteed positive. Furthermore, we have explored a recently proposed Sumca method (Jiang and Torabi 2020) in our simulation study. It was found that Sumca did not take any negative value in our case; however, it was found elsewhere that Sumca, too, can take negative values (Liu *et al.* 2022), but the probability that the Sumca estimator takes a negative value is very lower.

In our opinion (and this is also suggested by Jiang *et al.* 2018), log-MSPE is more convenient to estimate than MSPE in a way similar to that the log-likelihood is often easier to handle than the likelihood. Once a log-MSPE estimate is obtained, it can be easily converted to a MSPE estimate, which is guaranteed positive—one never has to worry that the MSPE estimate is negative; hence, for example, its square root cannot be taken, and used as a standard error of prediction. In particular, the practice of log-MSPE estimation should be encouraged in SAE.

Supplementary Materials

Technical derivations and expressions.

Acknowledgements

The research of Jianling Wang and Yihui Luan is supported by the National Science Foundation of China Grant No. 11971264 and the National Key R&D Program of China, Grant No. 2018YFA0703900. The research of Thuan Nguyen and Jiming Jiang is partially supported by the National Science Foundation of the United States grants DMS-1914760 and DMS-1914465, respectively. The authors are grateful to an Associate Editor and two referees for their constructive comments that have helped improve the manuscript.

References

Brooks, S. P., Morgan, B. J. T., Ridout, M. S., and Pack, S. E. (1997), Finite mixture models for proportions, *Biometrics* 53, 1097–1115.

Chen, S. and Lahiri, P. (2011), On the estimation of mean squared prediction error in small area estimation, *Calcutta Statist. Assoc. Bull.* 63, 249–252.

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

Jiang, J. (1998), Consistent estimators in generalized linear mixed models, *J. Amer. Statist. Assoc.* 93, 720-729.

Jiang, J. (2003), Empirical best prediction for small-area inference based on generalized linear mixed models, *J. Statist. Planning Inference* 111, 117–127.

Jiang, J., and Lahiri, P. (2001), Empirical best prediction for small area inference with binary data, *Ann. Inst. Statist. Math.* 53, 217–243.

Jiang, J., Lahiri, P., and Nguyen, T. (2018), A Monte-Carlo jackknife for small area estimation after model selection, *Ann. Math. Sci. Appl.* 3, 405–438.

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. and Nguyen, T. (2021), *Linear and Generalized Linear Mixed Models and Their Applications*, 2nd ed., Springer, New York.

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

Jiang, J. and Torabi, M. (2020), Sumca: Simple, unified, Monte-Carlo assisted approach to second-order unbiased MSPE estimation, *J. Roy. Statist. Soc. Ser. B* 82, 467–485

Jiang, J. and Zhang, W. (2001), Robust estimation in generalized linear mixed models, *Biometrika* 88, 753–765.

Liu, Y., Liu, X., Pan, Y., Jiang, J., and Xiao, P. (2022), A comprehensive comparison of finite sample performance of various MSPE estimators and associated prediction intervals in small area estimation, *J. Statist. Comp. Simulation*, revised.

Liu, X., Ma, H. and Jiang, J. (2022a), Assessing uncertainty in small area estimation involving model misspecification: A one-bring-one route, *Sci. China Math.*, in press.

Liu, X., Ma, H. and Jiang, J. (2022b), That Prasad-Rao is robust: Estimation of mean squared prediction error of observed best predictor under potential model misspecification, *Stat. Sin.*, in press.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed., Chapman and Hall, London.

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 ed., Wiley, New York.

Research Center for Mathematics and Interdisciplinary Sciences, Frontiers Science Center for Nonlinear Expectations (Ministry of Education), Shandong University, Qingdao, China.

E-mail: 1161407990@qq.com

OHSU/PSU School of Public Health, Oregon Health and Science University, Portland, OR, USA.

E-mail: nguythua@ohsu.edu

Research Center for Mathematics and Interdisciplinary Sciences, Frontiers Science Center for

Nonlinear Expectations (Ministry of Education), Shandong University, Qingdao, China.

E-mail: yhluan@sdu.edu.cn

Department of Statistics, University of California, Davis, USA.

E-mail: jimjiang@ucdavis.edu