Statistica Sinica Preprint No: SS-2022-0043	
Title	On Estimation of the Logarithm of the Mean Squared Prediction Error of A Mixed-effect Predictor
Manuscript ID	SS-2022-0043
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0043

Complete List of Authors	Jianling Wang, Thuan Nguyen, Yihui Luan and Jiming Jiang
Corresponding Authors	Jiming Jiang
E-mails	jimjiang@ucdavis.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

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) is 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. However, this is difficult, especially if the estimator needs to be positive, or at least nonnegative. In fact, very few MSPE estimators are both second-order unbiased and guaranteed to be positive. We consider an alternative, easier approach of estimating the logarithm of the MSPE (log-MSPE), thus avoiding the positivity problem. We derive a second-order unbiased estimator of the log-MSPE using the Prasad–Rao linearization method. The results of empirical studies demonstrate the superiority of the proposed log-MSPE estimator over a naive log-MSPE estimator and an existing method, known as McJack. Lastly, we demonstrate the proposed method by applying it to real data.

Key words and phrases: bias-correction, log-MSPE, mixed effects, second-order

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

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 et al. (2022a,b)) for some recent advances. However, this is difficult, especially if the MSPE estimator needs to be positive, or at least nonnegative. In fact, with very few exceptions (Prasad and Rao (1990),Chen and Lahiri (2011)), existing second-order unbiased MSPE estimators are not both second-order unbiased and guaranteed to be positive; see Jiang et al. (2018, p.408) for a 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, because the resulting second-order estimator is no longer guaranteed to be positive. One option is to modi-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

fy the value of the MSPE estimator when it is negative, for example, by truncating the estimator at zero, but this destroys the second-order unbiasedness. Jiang et al. (2018) use the work of Hall and Maiti (2006) as an example to illustrate this dilemma. Intuitively, this is analogous to the problem of trying to cover two ants, both moving fast in random directions, with two fingers of the same hand, which is relatively difficult. However, the task becomes much easier if we use one finger and there is just one ant, no matter how fast and randomly it moves.

Jiang et al. (2018) propose estimating the logarithm of the MSPE (log-MSPE), instead of the MSPE itself. The log-MSPE is a simple one-to-one transformation of the MSPE. Thus, we can easily convert a log-MSPE estimator to an MSPE estimator by taking the exponential. Furthermore, reporting the log-MSPE results often saves space, there are advantages in terms of hypothesis testing, and there is a linear association between the logarithm and the square root of the MSPE. Most importantly, positivity is not a concern, because the exponential of a log-MSPE estimator is always positive. Jiang et al. (2018) further propose a Monte Carlo (MC) jackknife method for estimating the log-MSPE, called McJack, showing that their method produces a second-order unbiased estimator of the log-MSPE.

In the SAE literature, two standard methods are used to produce a

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

second-order unbiased MSPE estimator, namely, the Prasad–Rao linearization method (Prasad and Rao (1990)) and the resampling method (e.g., Jiang et al. (2002),Hall and Maiti (2006)); see also Rao and Molina (2015). The McJack method is a resampling method. The main objective of our study is to develop a class of second-order unbiased estimators of the log-MSPE using the linearization method, and to demonstrate its advantages over 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 (GLMMs; e.g., Jiang and Nguyen (2021)). We present our simulation results in Section 4, and discuss a real-data example in Section 5. Section 6 concludes the paper.

2. Second-order unbiased log-MSPE estimator

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

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

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 following 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 definitions of O_{P} and 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 $a(\cdot), b(\cdot)$, and $c(\cdot)$ are continuous and may depend on m, but $a(\psi)$, $b(\psi)$, and $c(\psi)$ are bounded and $a(\psi)$ has a positive lower bound for ψ that satisfies (2.2)–(2.4). Note that (2.4) is a result of (iii) under regularity conditions. From the Taylor series expansion, we have

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

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

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 nonrandom. 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, from (2.7), we have

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

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

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

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

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

3. EBP based on GLMMs

In the context of SAE with discrete or categorical responses, Jiang (2003) proposes an EBP method based on a GLMMs. The method assumes that, conditional on the random-effect vectors $v_i = (v_{ij})_{1\leq j\leq n_i}$, for $1 \leq i \leq m$, the responses y_{ij}, for $1 \leq j \leq n_i$, are independent, with the conditional pmf, or pdf, given by

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

where $r(\cdot)$, and $s(\cdot,\cdot)$ are functions associated with the exponential family (McCullagh and Nelder (1989, ch.2)), ϕ is a dispersion parameter, which in some cases is known, and 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 a response

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

(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 the case of a canonical link, one has $\xi_{ij} = \eta_{ij}$. Finally, $v_1, \ldots, v_m$ 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 predicting a possibly nonlinear mixed effect in the form

$$\zeta \;=\; \zeta(\beta, v_S)\,, \tag{3.1}$$

where S is a subset of $\{1, \ldots, m\}$ 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 a minimum MSPE is given by

$$\begin{aligned}\tilde{\zeta} &= \frac{\int \zeta(\beta, v_S)\exp(\phi^{-1}\sum_{i\in S} s_i(\beta, v_i))\prod_{i\in S} f_\nu(v_i)\prod_{i\in S} dv_i}{\prod_{i\in S}\int \exp(\phi^{-1}s_i(\beta, v))f_\nu(v)dv} \\ &\equiv u(y_S, \psi), \end{aligned} \tag{3.2}$$

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 in (3.2) may be evaluated using numerical integration or MC methods. For the unknown parameters ψ, in (3.2), Jiang (2003) suggests using method

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

of moments (MoM) estimators, which are consistent (Jiang (1998)). Let $\hat{\psi}$ denote the MoM estimator of ψ. If we replace ψ in (3.2) with $\hat{\psi}$, we obtain the empirical BP, or EBP,

$$\hat{\zeta} \;=\; u(y_S, \hat{\psi}). \tag{3.3}$$

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

$$\mathrm{E}(\hat{\psi} - \psi)(\hat{\psi} - \psi)' = m^{-1}V(\psi) + o(m^{-1}). \tag{3.4}$$

Then, from Jiang (2003), we have the following expression:

$$\mathrm{MSPE} \equiv \mathrm{MSPE}(\hat{\zeta}) = d(\psi) + m^{-1}e(\psi) + o(m^{-1}), \tag{3.5}$$

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

$$M(\psi) = \hat{M}, \tag{3.6}$$

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, we have $\mathrm{E}(\hat{M}) = O(1)$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

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, from the Taylor series expansion at ψ, that is the true parameter vector, under regularity conditions, we have

$$\hat{M} = M(\hat{\psi}) = M + A(\hat{\psi} - \psi) + O_{\mathrm{P}}(m^{-1}), \tag{3.7}$$

where $A = \partial M/\partial\psi'$. Here (3.7) implies the following asymptotic expansion:

$$\hat{\psi} - \psi = A^{-1}(\hat{M} - M) + O_{\mathrm{P}}(m^{-1}), \tag{3.8}$$

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

$$\mathrm{E}(\hat{\psi} - \psi)(\hat{\psi} - \psi)' = A^{-1}\mathrm{E}(\hat{M} - M)(\hat{M} - M)'(A^{-1})' + o(m^{-1}). \tag{3.9}$$

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

$$V(\psi) = mA^{-1}\mathrm{Var}(\hat{M})(A^{-1})'. \tag{3.10}$$

Thus, combining (3.5) and (3.10), we have

$$\mathrm{MSPE} = d(\psi) + b_3(\psi) + o(m^{-1}) = d(\psi) + o(1), \tag{3.11}$$

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 $a(\psi) = d(\psi)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

Now, define $\widetilde{\text{MSPE}} = d(\hat{\psi})$. From the definition of $d(\psi)$, condition (i) of Section 2 is satisfied (assuming nonsingularity). Furthermore, from (3.11), we have

$$\mathrm{E}(\widetilde{\text{MSPE}} - \text{MSPE}) \;=\; \mathrm{E}\{d(\hat{\psi}) - d(\psi)\} - b_3(\psi) + o(m^{-1}). \quad (3.12)$$

Then, from the Taylor series expansion at the true ψ and (3.8), it can be shown that

$$\begin{aligned} d(\hat{\psi}) - d(\psi) \;=\; & \frac{\partial d}{\partial \psi'}(\hat{\psi} - \psi) + \frac{1}{2}(\hat{M} - M)'(A^{-1})'\frac{\partial^2 d}{\partial \psi \partial \psi'}A^{-1}(\hat{M} - M) \\ & + o_{\mathrm{P}}(m^{-1}). \end{aligned} \quad (3.13)$$

We can expand (3.8) to obtain a further expansion (see the Supplementary Material):

$$\begin{aligned} \hat{\psi} - \psi \;=\; & A^{-1}(\hat{M} - M) \\ & -\frac{1}{2}A^{-1}\left[(\hat{M} - M)'(A^{-1})'B_k A^{-1}(\hat{M} - M)\right]_{1 \le k \le q} \\ & + o_{\mathrm{P}}(m^{-1}), \end{aligned} \quad (3.14)$$

where $B_k = \partial^2 M_k / \partial \psi \partial \psi'$. Combining (3.13) and (3.14), we have, under regularity conditions, that

$$\mathrm{E}\{d(\hat{\psi}) - d(\psi)\} = \frac{b_1(\psi) - b_2(\psi)}{2} + o(m^{-1}), \quad (3.15)$$

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

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

Combining (3.12) and (3.15), it follows that (2.3) holds with

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

Finally, from (3.11), (3.13), and (3.14), it can be shown that

$$\widetilde{\mathrm{MSPE}} - \mathrm{MSPE} = \frac{\partial d}{\partial \psi'} A^{-1}(\hat{M} - M) + O_{\mathrm{P}}(m^{-1}). \tag{3.16}$$

Here, (3.16) implies that, under regularity conditions, (2.4) holds with

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

In conclusion, the general result of Section 2 applies, with $\widetilde{\mathrm{MSPE}} = d(\hat{\psi})$, $a(\psi) = d(\psi)$, and $\bar{b}(\psi)$ and $\bar{c}(\psi)$ specified above and below (3.16), respectively.

The following expressions are computationally more convenient:

$$b_1(\psi) = \mathrm{tr}\left((A^{-1})' \frac{\partial^2 d}{\partial \psi \partial \psi'} A^{-1} \mathrm{Var}(\hat{M}) \right), \tag{3.17}$$

$$b_2(\psi) = \frac{\partial d}{\partial \psi'} A^{-1} \left[\mathrm{tr}\left((A^{-1})' B_k A^{-1} \mathrm{Var}(\hat{M}) \right) \right]_{1 \le k \le q}, \tag{3.18}$$

$$b_3(\psi) = \mathrm{tr}\left(A^{-1} \mathrm{Var}(\hat{M}) (A^{-1})' \mathrm{E}\left(\frac{\partial u}{\partial \psi} \frac{\partial u}{\partial \psi'} \right) \right), \tag{3.19}$$

where the expectation inside the trace is with respect to y_S in (3.2).

Computational/practical notes

1. In theory, $d(\psi) = \mathrm{MSPE}(\tilde{\zeta})$ should be positive for any ψ. However, 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

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

integration to evaluate $d(\psi)$. Then, owing to the integral approximations, $d(\hat{\psi})$ can occasionally take negative values. When the value of $d(\hat{\psi})$ is negative, we suggest evaluating it using an MC 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. The matrix A can occasionally be singular. In this case, we suggest using the Moore–Penrose generalized inverse of A in place of A^{-1}.

3. In some cases, there are known bounds for the value of the 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 one, and hence the log-MSPE is bounded by zero. Thus, the value of the log-MSPE estimate is taken as zero (hence, the MSPE estimate is equal to one) when it is greater than zero.

4. Example and simulation

Consider a mixed logistic model for SAE (e.g., Jiang and Lahiri (2001)). Suppose that, conditional on p_i, y_{ij} are independent *Bernoulli*, with $P(y_{ij} = 1|p_i) = p_i$, for $i = 1, \cdots, m$ and $j = 1, \cdots, k_i$. In addition, we have $\text{logit}(p_i) = \log(p_i/(1-p_i)) = \mu + v_i$, where μ is a known parameter. Furthermore, $v_1, \ldots, v_m$ are independent random effects. Two distributions of

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

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 $f(x|\sigma) = (2\sigma)^{-1}e^{-|x|/\sigma}$, for $-\infty < x < \infty$.

For simplicity, let $k_i = k > 1$, for $1 \le i \le m$. 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 a 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

$$\tilde{p}_i = e^{\mu}\frac{\mathrm{E}\{\exp((y_{i\cdot}+1)\sigma\xi - (k+1)\log(1+e^{\mu+\sigma\xi}))\}}{\mathrm{E}\{\exp(y_{i\cdot}\sigma\xi - k\log(1+e^{\mu+\sigma\xi}))\}} \equiv u(y_{i\cdot}, \psi), \quad (4.1)$$

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 $\tilde{p}_i$, replacing ψ with $\hat{\psi}$, the MoM estimator. The latter is the solution to (3.6) 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 $M_s(\psi) = \mathrm{E}\{h^s(\mu+\sigma\xi)\}$, for $s = 1, 2$ (Jiang (1998)). We have the following expression (see the Supplementary Material):

$$\begin{aligned} d(\psi) &= \mathrm{E}\{h^2(\mu+\sigma\xi)\} \\ &- \sum_{l=0}^k u^2(l,\psi)\binom{k}{l}\mathrm{E}\left\{\exp\left(l(\mu+\sigma\xi) - k\log\left(1+e^{\mu+\sigma\xi}\right)\right)\right\}, \quad (4.2) \end{aligned}$$

where $u(l, \psi)$ is $u(y_{i\cdot}, \psi)$ [see (4.1)] with $y_{i\cdot} = l$.

For notational simplicity, we write $h = h(\mu + \sigma\xi)$ when ψ is the true

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

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 S2.2 of the Supplementary Material, and those of the partial derivatives in (3.17)–(3.19) are given in Section S2.3 of the Supplementary Material. Note that, in this case, we have

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

where $s(a,b,w) = \exp(aw - b\log(1+e^w))$. All of the expectations are evaluated by means of numerical integration, using the integrate() function in R (lower bound $= -5$; upper bound $= 5$). Furthermore, in this case, the MSPE is naturally bounded by one, and hence the log-MSPE is bounded by zero. Thus, the value of the log-MSPE estimate is taken as zero when it is positive (see Note 3 at the end of the previous section).

Simulation studies are carried out to evaluate the performance of the proposed bias-corrected log-MSPE estimator given in (2.8). We compare

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

the proposed estimator with a naive log-MSPE estimator, which is simply $\log(\widetilde{\text{MSPE}})$, the first term on the right-hand side of (2.8). Consider predicting p_1 using the EBP. Here, we let $m = 25, 50, 100$ and $k_i = 4$, for $1 \leq i \leq m$. The true parameters are $\mu = -1.0$ and $\sigma = 2.0$. The MC sample size used to evaluate $d(\hat{\psi})$ when it is occasionally negative (see Note 1 at the end of Section 3) is $N_{\text{mc}} = 1,000$.

We consider the following performance measures:

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

(2) Percentage relative bias (%RB), given by $100 \times$ (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 from the simulation runs.

The results, based on $N_{\text{sim}} = 2,000$ simulation runs, are presented in Table 1, showing that the bias-corrected estimator, $\widehat{\log}(\text{MSPE})$, performs best 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 used to estimate the log-MSPE. Because the McJack method is computationally intensive, and the com-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

Table 1: **Comparison with the naive log-MSPE estimator**

Case	Sample Size	Simulated log-MSPE	$\log(\widetilde{\text{MSPE}})$ Bias	%RB	CV	$\widehat{\log}(\text{MSPE})$ 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

putational burden increases quickly with m, the comparison is limited to the case of $m = 25$, and $k = 4$. In addition to the above performance measures, we also consider the average computing time (ACT; in seconds) per simulation run. Because the McJack method depends on the MC sample size used to evaluate the expectations, we consider two MC sample sizes, $K_{\text{mc}} = 50, 100$. Owing to the computational intensity of the McJack method, we set $N_{\text{sim}} = 500$ (instead of $N_{\text{sim}} = 2,000$, as in the previous case). The results are reported in Table 2, in which the results for $\widehat{\log}(\text{MSPE})$, with the exception of those for the ACT, are copied from

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

Table 2: **Comparison with the McJack method**

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

Table 1.

The results show that $\widehat{\log}$(MSPE) performs better in terms of both %RB and CV, although the results are comparable. The biggest difference is in terms of the computational efficiency, where $\widehat{\log}$(MSPE) performs significantly better. For example, in case (a), the ACT of the McJack method 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, owing to the computational intensity, we consider only the case of $m = 25$. When m is larger, the computing time needed for the McJack method may become impractical. This leaves $\widehat{\log}$(MSPE) as the only feasible method capable of producing a second-order unbiased log-MSPE estimator when m is large.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

5. A real-data example

Brooks et al. (1997) present six data sets on recording fetal mortality in mouse litters. As an application, we consider the HS2 data set 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 using a GLMM (see also Jiang and Nguyen (2021, sec.4.4.1)). Let y_{ij}, for $i = 1, ..., m$, and $j = 1, \ldots, k_i$ 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} are assumed to satisfy the same mixed logistic model with normally distributed random effects described at the beginning of Section 4. We also considered the mixed logistic model with Laplacian random effects, 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_i are not equal). Thus, the definition of $\hat{M}_s$, for $s = 1, 2$ differs from 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 $\hat{M}_2 = \{\sum_{i=1}^m k_i(k_i - 1)\}^{-1} \sum_{i=1}^m (y_{i\cdot}^2 - y_{i\cdot})$. Following Jiang and Nguyen (2021, sec.4.4.1), the MoM estimates are $\hat{\mu} = -2.276$ and $\hat{\sigma} = 0.644$. The expression for $\mathrm{Var}(\hat{M})$ in this case and additional expressions in terms of the current data structure are given in Section S3 of the Supplementary

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

Material.

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, depend only on the values of $\hat{\mu}, \hat{\sigma}, k_i$, and $y_{i\cdot}$. In this case, the McJack estimates are computationally intensive ($m = 1,328$ in this case); see the 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 using numerical integration. The results, including the EBP and corresponding square root of the MSPE (RMSPE) estimate, obtained from a simple transformation of 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.

The results show that the EBP decreases as k_i increases, and increases as $y_{i\cdot}$ increases. Although both trends can be shown theoretically, there are also intuitive explanations. Recall that the EBP predicts the conditional probability that the implant is dead, given the observed count, $y_{i\cdot}$. For example, take consider $k_i = 7$. If $y_{i\cdot}$ is zero, that is, no implant is dead, the predicted probability of death is 0.084. If $y_{i\cdot} = 1$, that is, one implant

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

Table 3: **Analysis of mice mortality data:** for each number of implants, the first row shows the observed number of cases, and the second row show the 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		

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

is dead, one would expect the conditional probability of death to increase; this is indeed the case, because 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, we expect more death; therefore, the probability of exactly one death should decrease.

Note that the RMSPE depends only 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, $y_{i\cdot}$, for $i = 1, \ldots, m$, are independent and identically distributed, with a distribution that depends only 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 size. As k_i increases, more information is available for better prediction; as a result, the MSPE should decrease.

6. Conclusion

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) nonlinear mixed effect using the EBP under a GLMM. We demonstrate the superiority of our method over a naive predictor and the McJack method, especially in terms of the computational efficiency. We use a real-data

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

example to illustrate the practical relevance of our method.

The computational disadvantage of the McJack method makes it difficult to evaluate its performance in large-scale simulation studies, which need many simulation runs 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 the McJack method, and sometimes more accurate.

The current approach is similar to the Prasad–Rao linearization method for estimating the MSPE (Prasad and Rao (1990)). Thus, it does not result in any simplification in terms of the analytic derivations compared with the Prasad–Rao method. However, there is potentially a middle ground between the analytically tedious Prasad–Rao method and the computationally intensive McJack method. Recently, Jiang and Totabi (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: it uses the linearization method to obtain the leading term of the MSPE estimator, and MC method to obtain a bias correction to achieve the second-order unbiasedness. The linearization is (much) simpler to derive than that of the Prasad–Rao method (because one does not need to

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

achieve the second-order unbiasedness for the leading term), and the MC bias correction is computationally much faster than those of the McJack or double bootstrap methods (Hall and Maiti (2006)). In future work, we shall explore extending the Sumca method to the log-MSPE estimation.

As mentioned in Section 1, most existing second-order unbiased MSPE estimators may take negative values. In particular, Liu et al. (2022b) propose 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 theoretical proof that this estimator is guaranteed to be positive. Furthermore, we explored the recently proposed Sumca method (Jiang and Totabi (2020)) in our simulation study, finding that it did not take any negative values in our case. Although Liu et al. (2022b) show that it can take negative values, the probability of this occurring is very low.

In our opinion (and this is also suggested by Jiang et al. (2018), the log-MSPE is more convenient to estimate than the MSPE, similarly to show the log-likelihood is often easier to handle than the likelihood. Once we have a log-MSPE estimate, it can be converted easily into an MSPE estimate, which is guaranteed to be positive. In particular, log-MSPE estimation should be encouraged in SAE.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

Supplementary Material

Technical derivations and expressions are provided in the online Supplementary Material.

Acknowledgments

The research of Jianling Wang and Yihui Luan was 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 was partially supported by the National Science Foundation of the United States grants DMS-1914760 and DMS-1914465, respectively. The authors are grateful to the associate editor and two referees for their constructive comments and suggestions.

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.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

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. (2010), *Large sample techniques for statistics*, Springer, New York.

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.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

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

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0043

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

Statistica Sinica