CIRJE-F-934

On Conditional Mean Squared Errors of Empirical Bayes Estimators in Mixed Models with Application to Small Area Estimation

Shonosuke Sugasawa
The University of Tokyo

Tatsuya Kubokawa
The University of Tokyo

June 2014

On Conditional Mean Squared Errors of Empirical Bayes Estimators in Mixed Models with Application to Small Area Estimation

Shonosuke Sugasawa* and Tatsuya Kubokawa†

University of Tokyo

June 5, 2014

Abstract

This paper is concerned with the prediction of the conditional mean which involves the fixed and random effects based on the natural exponential family with a quadratic variance function. The best predictor is interpreted as the Bayes estimator in the Bayesian context, and the empirical Bayes estimator (EB) is useful for small area estimation in the sense of increasing precision of prediction for small area means. When data of the small area of interest are observed and one wants to know the prediction error of the EB based on the data, the conditional mean squared error (cMSE) given the data is used instead of the conventional unconditional MSE. The difference between the two kinds of MSEs is small and appears in the second-order terms in the classical normal theory mixed model. However, it is shown that the difference appears in the first-order or leading terms for distributions far from normality. Especially, the leading term in the cMSE is a quadratic concave function of the direct estimate in the small area for the binomial-beta mixed model, and an increasing function for the the Poisson-gamma mixed model, while the leading terms in the unconditional MSEs are constants for the two mixed models. Second-order unbiased estimators of the cMSE are provided in two ways based on the analytical and parametric bootstrap methods. Finally, the performances of the EB and the estimator of cMSE are examined through simulation and empirical studies.

Key words and phrases: Binomial-beta mixture model, conditional mean squared error, Fay-Herriot model, mixed model, natural exponential family with quadratic variance function, Poisson-gamma mixture model, random effect, small area estimation.

1 Introduction

The empirical best linear unbiased predictors (EBLUP), which are empirical Bayes estimators (EB) in the Bayesian context, have been recognized to give reliable small area estimates in

*Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN, E-Mail: shonosuke622@gmail.com

†Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN, E-Mail: tatsuya@e.u-tokyo.ac.jp

the normal linear mixed models. The unconditional mean squared errors (MSE) have been widely used as measure of uncertainty of EB, and their asymptotic approximations and their approximated unbiased estimators have been studied in a lot of papers. For example, see Prasad and Rao (1990), Ghosh and Rao (1994), Rao (2003), Datta, Rao and Smith (2005) and Hall and Maiti (2006). When data of the small area of interest are observed, the practitioners want to know how large prediction errors the EB estimates based on the observed data have. Concerning this issue, the conventional unconditional MSEs do not give us appropriate estimation errors, since it is an integrated measure. Booth and Hobert (1998) suggested the use of the conditional MSE given the data of the small area of interest, and Datta, Kubokawa, Molina and Rao (2011) and Torabi and Rao (2013) derived second-order unbiased estimators of the conditional MSE in the Fay-Herriot and nested error regression models which are normal linear mixed models. As pointed out in both papers, the difference between the unconditional and conditional MSEs is small in the normal linear mixed models, since it appears in the second-order terms. In the generalized linear mixed models, however, Booth and Hobert (1998) showed that the difference is significant for distributions far from normality, namely, it appears in the first-order or leading terms. To give a second-order unbiased estimator of the conditional MSE, they used the Laplace approximation under the assumption that the sample size from the small area is large. It should be regrettable that this assumption is against the situation in the small area estimation. In this paper, we consider mixed models based on natural exponential families with quadratic variance functions (NEF-QVF), and derive second-order approximations of cMSE of EB and their second-order unbiased estimators as analytical methods under the assumption that the sample size from the small area is bounded. For the small area estimation based on NEF-QVF and the related studies, see Ghosh and Maiti (2004, 2008) and Kubokawa, Hasukawa and Takahashi (2014).

To explain more details, let us consider the mixed model that $(y_1, \theta_1), \ldots, (y_m, \theta_m)$ be mutually independent random variables such that for each i, the conditional distribution of y_i given θ_i depends on θ_i and an unknown parameter $\boldsymbol{\eta}$, and the marginal distribution of θ_i depends on $\boldsymbol{\eta}$. It is assumed that we want to predict a scalar quantity $\xi_i(\theta_i, \boldsymbol{\eta})$ based on the observations $y_1, \ldots, y_m$. Then the posterior mean of $\xi_i(\theta_i, \boldsymbol{\eta})$ is given by $\widehat{\xi}_i(y_i, \boldsymbol{\eta}) = E[\xi_i(\theta_i, \boldsymbol{\eta})|y_i]$. When the parameter is consistently estimated with $\widehat{\boldsymbol{\eta}} = \widehat{\boldsymbol{\eta}}(y_1, \ldots, y_m)$, one can estimate $\xi_i(\theta_i, \boldsymbol{\eta})$ with $\widehat{\xi}_i = \widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}})$. This estimator is evaluated in two ways: the unconditional and conditional MSEs, given by

$$\begin{aligned}
\mathrm{MSE}(\boldsymbol{\eta}, \widehat{\xi}_i) =& E_{\boldsymbol{\eta}}[\{\widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}}) - \xi_i(\theta_i, \boldsymbol{\eta})\}^2], \\
\mathrm{cMSE}(\boldsymbol{\eta}, \widehat{\xi}_i) =& E_{\boldsymbol{\eta}}[\{\widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}}) - \xi_i(\theta_i, \boldsymbol{\eta})\}^2|y_i],
\end{aligned}$$

which are denoted by MSE and cMSE, respectively. When $\boldsymbol{\eta}$ is known, the best predictors of $\xi_i(\theta_i, \boldsymbol{\eta})$ in terms of the two kinds of MSEs are the posterior mean $\widehat{\xi}_i(y_i, \boldsymbol{\eta})$, which is the Bayes estimator. Thus, the plug-in estimator $\widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}})$ corresponds to the empirical Bayes estimator (EB). Booth and Hobert (1998) showed that the difference between the MSE and cMSE appears in the second-order terms in the normal linear mixed models. Also they demonstrated that the difference is significant for distributions far from the normality in the generalized linear mixed models, where the leading terms in the cMSE are approximated by the Laplace approximation under the assumption that the sample sizes from small areas are large.

In this paper, we revisit the same problem and investigate more clearly the difference between the cMSE and the MSE. As a specific model, we treat the mixed model based on the natural exponential families with quadratic variance functions (NEF-QVF). In this model, the leading terms in the cMSE are expressed in explicit forms without assuming that the sample sizes from small areas tend to infinity.

In Section 2, we begin by approximating the cMSE of EB in the general mixed models under some suitable conditions on $\widehat{\boldsymbol{\eta}}$ and $\widehat{\xi}_i(y_i, \boldsymbol{\eta})$. The leading term in the cMSE is given by the posterior variance $Var(\xi_i(\theta_i, \boldsymbol{\eta})|y_i)$, which is of order $O(1)$. Second-order unbiased estimators of cMSE are provided in two ways of the analytical and parametric bootstrap methods.

As an application of the general results in Section 2, we treat the mixed models based on the NEF-QVF in Section 3. The unconditional MSE was derived by Ghosh and Maiti (2004), who used the estimating equations suggested in Godambe and Thompson (1989) for estimating the unknown parameter $\boldsymbol{\eta}$. We employ the same methods and techniques as used in Ghosh and Maiti (2004). The feature of the NEF-QVF is that the conditional variance of y_i given θ_i is a quadratic function of the mean $\xi_i = E[y_i|\theta_i]$, namely,

$$Var(y_i|\theta_i) = Q(\xi_i)/n_i,$$

where n_i is a known constant, and $Q(x) = v_0 + v_1 x + v_2 x^2$ for constants v_0, v_1 and v_2, which are not simultaneously zero. For the normal, Poisson and binomial distributions, (v_0, v_1, v_2) corresponds to $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 1, -1)$, respectively. Then, it is demonstrated that the leading terms in the cMSE are expressed in explicit forms of

$$Var(\xi_i|y_i) = Q(\widehat{\xi}_i)/(n_i + \nu - v_2),$$

where $\widehat{\xi}_i = (n_i y_i + \nu m_i)/(n_i + \nu)$ for unknown model parameters ν and m_i. This shows that the leading term is a constant for the normal distribution, while it is an increasing function of y_i for the Poisson distribution and a quadratic concave function of y_i for the binomial distribution.

In Section 3, we provide a second-order approximation of the cMSE without assuming that n_i tends to infinity. We also derive an analytical and closed form of a second-order unbiased estimator of cMSE. In the generalized linear mixed models, it is hard to derive an analytical estimator with a closed form for the cMSE, and Booth and Hobert (1998) suggested an estimator using the parametric bootstrap method. The result in Section 3 means that it is possible in the mixed models based on NEF-QVF, however. Some examples are illustrated for the Fay-Herriot, the Poisson-gamma mixture and the binomial-beta mixture models.

The simulation and empirical studies are reported in Section 4. For the empirical studies, we treat two data sets. One is the Stomach Cancer Mortality Data in Saitama prefecture in Japan, and we apply the Poisson-gamma mixture model. The other is the Infant Mortality Data Before World War II in Ishikawa prefecture in Japan, and we use the binomial-beta mixture model since the mortality rate is distributed around $p = 0.2$. Since cMSE depends on the data of the area of interest, the estimates of cMSE are more variable than those of MSE. For some areas, cMSE gives much higher risks than MSE, and we should note that the conventional MSE seems to under-estimate a prediction error of the EB estimate for given data of the area. Thus, we recommend to provide the estimates of cMSE as well as the estimates of MSE. Finally, the concluding remarks are given in Section 5, and the technical proofs are given in the Appendix.

2 Conditional MSE of Empirical Bayes Estimator in General Mixed Models

Let $\boldsymbol{y} = (y_1, \ldots, y_m)^t$ be a vector of observable random variables, and let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^t$ be a vector of unobservable random variables. Let $\boldsymbol{\eta}$ be a q-dimensional vector of unknown parameters. In this paper, we treat continuous or discrete cases for y_i and $\boldsymbol{\theta}$. The conditional probability density (or mass) function of y_i given $(\theta_i, \boldsymbol{\eta})$ is denoted by $f(y_i|\theta_i, \boldsymbol{\eta})$, and the conditional probability density (or mass) function of θ_i given $\boldsymbol{\eta}$ is denoted by $\pi(\theta_i|\boldsymbol{\eta})$, namely,

$$\begin{aligned} y_i|(\theta_i, \boldsymbol{\eta}) \sim& f(y_i|\theta_i, \boldsymbol{\eta}) \\ \theta_i|\boldsymbol{\eta} \sim& \pi(\theta_i|\boldsymbol{\eta}) \end{aligned} \qquad i = 1, \ldots, m. \tag{1}$$

This expresses the general parametric mixed models. Since it can be interpreted as a Bayesian model, we here use the terminology used in Bayes statistics. In the continuous case, the marginal density function of y_i for given $\boldsymbol{\eta}$ and the conditional (or posterior) density function of θ_i given $(y_i, \boldsymbol{\eta})$ are given by

$$\begin{aligned} m_\pi(y_i|\boldsymbol{\eta}) =& \int f(y_i|\theta_i, \boldsymbol{\eta})\pi(\theta_i|\boldsymbol{\eta})\mathrm{d}\theta_i \\ \pi(\theta_i|y_i, \boldsymbol{\eta}) =& f(y_i|\theta_i, \boldsymbol{\eta})\pi(\theta_i|\boldsymbol{\eta})/m_\pi(y_i|\boldsymbol{\eta}) \end{aligned} \qquad i = 1, \ldots, m, \tag{2}$$

and we use the same notations in the discrete case. Then, for $i = 1, \ldots, m$, we consider the problem of predicting a scalar quantity $\xi_i(\theta_i, \boldsymbol{\eta})$ of each small area.

When $\xi_i(\theta_i, \boldsymbol{\eta})$ is predicted with $\widehat{\xi}_i = \widehat{\xi}_i(\boldsymbol{y})$, the predictor $\widehat{\xi}_i$ can be evaluated with the unconditional and conditional MSEs, described as

$$\begin{aligned} \mathrm{MSE}(\boldsymbol{\eta}, \widehat{\xi}_i) =& E\left[\left\{\widehat{\xi}_i - \xi_i(\theta_i, \boldsymbol{\eta})\right\}^2\right], \\ \mathrm{cMSE}(\boldsymbol{\eta}, \widehat{\xi}_i|y_i) =& E\left[\left\{\widehat{\xi}_i - \xi_i(\theta_i, \boldsymbol{\eta})\right\}^2|y_i\right], \end{aligned}$$

which are denoted by MSE and cMSE, respectively. The best predictors of $\xi_i(\theta_i, \boldsymbol{\eta})$ in terms of the two kinds of MSEs are the conditional mean given by

$$\widehat{\xi}_i(y_i, \boldsymbol{\eta}) = E\left[\xi_i(\theta_i, \boldsymbol{\eta})|y_i\right],$$

which is the Bayes estimator in the Bayesian context. Since $\boldsymbol{\eta}$ is unknown, we need to estimate $\boldsymbol{\eta}$ from observations $y_1, \ldots, y_m$. Substituting an estimator $\widehat{\boldsymbol{\eta}}$ into $\widehat{\xi}_i(y_i, \boldsymbol{\eta})$ results in the empirical Bayes (EB) estimator $\widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}})$.

In this paper, we focus on asymptotic evaluations of the cMSE. To this end, we assume the following conditions on the estimator $\widehat{\boldsymbol{\eta}}$ and the predictor $\widehat{\xi}_i(y_i, \boldsymbol{\eta})$ for large m:

Assumption 1.

(A.1) *The dimension q of $\boldsymbol{\eta}$ is bounded and the estimator $\widehat{\boldsymbol{\eta}}$ satisfies that $(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})|y_i = \boldsymbol{O}_p(m^{-1/2})$ and $E[\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}|y_i] = O_p(m^{-1})$ for $i = 1, \ldots, m$.*

(A.2) *For* $i = 1, \ldots, m$, $\xi_i(\theta_i, \boldsymbol{\eta}) = O_p(1)$ *and* $\widehat{\xi}_i(y_i, \boldsymbol{\eta}) = O_p(1)$. *The estimator* $\widehat{\xi}_i(y_i, \boldsymbol{\eta})$ *is continuously differentiable with respect to* $\boldsymbol{\eta}$, *and*

$$\partial\widehat{\xi}_i(y_i, \boldsymbol{\eta})/\partial\boldsymbol{\eta} = O_p(1).$$

Under conditions (A1) and (A2), we get a second-order approximation of cMSE of $\widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}})$. Let

$$\begin{aligned} T_{1i}(y_i, \boldsymbol{\eta}) =& Var(\xi_i(\theta_i, \boldsymbol{\eta})|y_i), & (3)\\ T_{2i}(y_i, \boldsymbol{\eta}) =& E\Big[\Big\{(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t \frac{\partial\widehat{\xi}_i(y_i, \boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\Big\}^2\Big|y_i\Big], & (4) \end{aligned}$$

where $T_{1i}(y_i, \boldsymbol{\eta})$ is the conditional or posterior variance of $\xi_i(\theta_i, \boldsymbol{\eta})$. It is noted that $T_{1i}(y_i, \boldsymbol{\eta}) = O_p(1)$ and $T_{2i}(y_i, \boldsymbol{\eta}) = O_p(m^{-1})$ under Assumption 1.

Theorem 1. *Under assumption 1, the conditional MSE of* $\widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}})$ *is approximated as*

$$\text{cMSE}(\boldsymbol{\eta}, \widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}})|y_i) = T_{1i}(y_i, \boldsymbol{\eta}) + T_{2i}(y_i, \boldsymbol{\eta}) + o_p(m^{-1}). \tag{5}$$

Proof. Since $E[\xi_i - \widehat{\xi}_i(y_i, \boldsymbol{\eta})|y_i] = 0$, it is observed that

$$\begin{aligned} \text{cMSE}(\boldsymbol{\eta}, \widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}})|y_i) =& E[\{\xi_i(\theta_i, \boldsymbol{\eta}) - \widehat{\xi}_i(y_i, \boldsymbol{\eta}) + \widehat{\xi}_i(y_i, \boldsymbol{\eta}) - \widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}})\}^2|y_i]\\ =& E[\{\xi_i(\theta_i, \boldsymbol{\eta}) - \widehat{\xi}_i(y_i, \boldsymbol{\eta})\}^2|y_i] + E[\{\widehat{\xi}_i(y_i, \boldsymbol{\eta}) - \widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}})\}^2|y_i], \quad (6) \end{aligned}$$

and that $E[\{\xi_i(\theta_i, \boldsymbol{\eta}) - \widehat{\xi}_i(y_i, \boldsymbol{\eta})\}^2|y_i] = Var(\xi(\theta_i, \boldsymbol{\eta})|y_i) = T_{1i}(y_i, \boldsymbol{\eta})$. It is noted that

$$\widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}}) = \widehat{\xi}_i(y_i, \boldsymbol{\eta}) + \Big(\frac{\partial\widehat{\xi}_i(y_i, \boldsymbol{\eta}^*)}{\partial\boldsymbol{\eta}}\Big)^t(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}),$$

where $\boldsymbol{\eta}^*$ is between $\boldsymbol{\eta}$ and $\widehat{\boldsymbol{\eta}}$. Thus, we obtain

$$E[\{\widehat{\xi}_i(y_i, \boldsymbol{\eta}) - \widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}})\}^2|y_i] = E\Big[\Big\{(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t \frac{\partial\widehat{\xi}_i(y_i, \boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\Big\}^2\Big|y_i\Big] + o_p(m^{-1}),$$

which shows Theorem 1. □

We next derive second-order unbiased estimators of T_1 and T_2, which result in a second-order unbiased estimator of cMSE. As seen from Theorem 1, the order of $T_{2i}(y_i, \boldsymbol{\eta})$ is $O_p(m^{-1})$, so that we can estimate $T_{2i}(y_i, \boldsymbol{\eta})$ by $T_{2i}(y_i, \widehat{\boldsymbol{\eta}})$ unbiasedly up to second-order. For estimation of $T_{1i}(y_i, \boldsymbol{\eta})$, the naive estimator $T_{1i}(y_i, \widehat{\boldsymbol{\eta}})$ has a second-order bias because $T_{1i}(y_i, \boldsymbol{\eta}) = O_p(1)$. It is observed that

$$E[T_{1i}(y_i, \widehat{\boldsymbol{\eta}})|y_i] = T_{1i}(y_i, \boldsymbol{\eta}) + T_{11i}(y_i, \boldsymbol{\eta}) + T_{12i}(y_i, \boldsymbol{\eta}) + o_p(m^{-1}), \tag{7}$$

where

$$T_{11i}(y_i, \boldsymbol{\eta}) = \Big(\frac{\partial T_{1i}(y_i, \boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\Big)^t E[(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})|y_i] \tag{8}$$

and

$$T_{12i}(y_i,\boldsymbol{\eta}) = \frac{1}{2}\mathrm{tr}\Big[\Big(\frac{\partial^2 T_{1i}(y_i,\boldsymbol{\eta})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^t}\Big)E\left[(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})^t|y_i\right]\Big]. \tag{9}$$

It is noted that $T_{11i}(y_i,\boldsymbol{\eta}) = O_p(m^{-1})$ and $T_{12i}(y_i,\boldsymbol{\eta}) = O_p(m^{-1})$ under Assumption 1.

[**Analytical method**] It follows from (7) that a second-order unbiased estimator of cMSE is given by

$$\widehat{\mathrm{cMSE}}_i(\widehat{\xi}_i(y_i,\widehat{\boldsymbol{\eta}})) = T_{1i}(y_i,\widehat{\boldsymbol{\eta}}) - T_{11i}(y_i,\widehat{\boldsymbol{\eta}}) - T_{12i}(y_i,\widehat{\boldsymbol{\eta}}) + T_{2i}(y_i,\widehat{\boldsymbol{\eta}}). \tag{10}$$

Theorem 2. *Under Assumption 1, the estimator* (10) *is a second-order unbiased estimator of cMSE, namely*

$$E\big[\widehat{\mathrm{cMSE}}_i(\widehat{\xi}_i(y_i,\widehat{\boldsymbol{\eta}}))|y_i\big] = \mathrm{cMSE}(\boldsymbol{\eta},\widehat{\xi}_i(y_i,\widehat{\boldsymbol{\eta}})|y_i) + o_p(m^{-1}).$$

As explained in Section 3, in the mixed model based on NEF-QVF, we can provide analytical expressions for T_{11i} and T_{12i}, whereby we obtain a second-order unbiased estimator in a closed form. In general, however, it is hard to obtain analytical expressions for T_{11i} and T_{12i}. In this case, as given below, the parametric bootstrap method helps us obtain a feasible second-order unbiased estimator of cMSE.

[**Parametric bootstrap method**] Since y_i is fixed, a bootstrap sample is generated from

$$y_j^*|(\theta_j^*,\widehat{\boldsymbol{\eta}}) \sim f(y_j^*|\theta_j^*,\widehat{\boldsymbol{\eta}}) \qquad j\neq i,\ j=1,\dots,m,$$

where θ_j^*'s are mutually independently distributed as $\theta_j^*|\widehat{\boldsymbol{\eta}} \sim \pi(\theta_j^*|\widehat{\boldsymbol{\eta}})$. Noting that y_i is fixed, we construct the estimator $\widehat{\boldsymbol{\eta}}_{(i)}^*$ from the bootstrap sample

$$y_1^*,\dots,y_{i-1}^*,y_i,y_{i+1}^*,\dots,y_m^* \tag{11}$$

with the same technique as used to obtain the estimator $\widehat{\boldsymbol{\eta}}$. Let $E_*\left[\cdot|y_i\right]$ be the expectation with regard to the bootstrap sample (11). A second-order unbiased estimator of $T_{1i}(y_i,\boldsymbol{\eta})$ is given by

$$\overline{T}_{1i}(y_i,\widehat{\boldsymbol{\eta}}) = 2T_{1i}(y_i,\widehat{\boldsymbol{\eta}}) - E_*\left[T_{1i}(y_i,\widehat{\boldsymbol{\eta}}_{(i)}^*)|y_i\right].$$

Then, it can be verified that $E[\overline{T}_{1i}(y_i,\widehat{\boldsymbol{\eta}})|y_i] = T_{1i}(y_i,\boldsymbol{\eta}) + o_p(m^{-1})$. In fact, from (7), it is noted that

$$E[T_{1i}(y_i,\widehat{\boldsymbol{\eta}})|y_i] = T_{1i}(y_i,\boldsymbol{\eta}) + d_i(y_i,\boldsymbol{\eta}) + o_p(m^{-1}),$$

where $d_i(y_i,\boldsymbol{\eta}) = T_{11i}(y_i,\boldsymbol{\eta}) + T_{12i}(y_i,\boldsymbol{\eta})$. This implies that $E_*\left[T_{1i}(y_i,\widehat{\boldsymbol{\eta}}_{(i)}^*)|y_i\right] = T_{1i}(y_i,\widehat{\boldsymbol{\eta}}) + d_i(y_i,\widehat{\boldsymbol{\eta}}) + o_p(m^{-1})$. Since $d_i(y_i,\boldsymbol{\eta})$ is continuous in $\boldsymbol{\eta}$ and $d_i(y_i,\boldsymbol{\eta}) = O_p(m^{-1})$, one gets $E[\overline{T}_{1i}(y_i,\widehat{\boldsymbol{\eta}})|y_i] = T_{1i}(y_i,\boldsymbol{\eta}) + o_p(m^{-1})$.

For $T_{2i}(y_i,\boldsymbol{\eta})$, from (6), it is estimated via parametric bootstrap method as

$$T_{2i}^*(y_i,\widehat{\boldsymbol{\eta}}) = E^*\big[\{\widehat{\xi}_i^*(y_i,\widehat{\boldsymbol{\eta}}) - \widehat{\xi}_i^*(y_i,\widehat{\boldsymbol{\eta}}_{(i)}^*)\}^2|y_i\big].$$

It is noted that the estimator $T_{2i}^*(y_i,\widehat{\boldsymbol{\eta}})$ is always available although an analytical expression of $T_{2i}(y_i,\boldsymbol{\eta})$ is not necessarily available. Combining the above results yields the estimator

$$\widehat{\mathrm{cMSE}}_i^*(\widehat{\xi}_i(y_i,\widehat{\boldsymbol{\eta}})) = \overline{T}_{1i}(y_i,\widehat{\boldsymbol{\eta}}) + T_{2i}^*(y_i,\widehat{\boldsymbol{\eta}}). \tag{12}$$

Theorem 3. *Under Assumption 1, the estimator* (12) *is a second-order unbiased estimator of cMSE, namely*

$$E[\widehat{\text{cMSE}}_i^*|y_i] = \text{cMSE}(\boldsymbol{\eta}, \widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}})|y_i) + o_p(m^{-1}).$$

3 Applications to NEF-QVF

We now apply the results in the previous section to the mixed models based on natural exponential families with quadratic variance functions (NEF-QVF). The mixed models are used in context of small area estimation by Ghosh and Maiti (2004), who derived the second-order approximation and its unbiased estimator of the unconditional MSE for calibrating uncertainty of the empirical Bayes estimator. In this section, we handle an area level model with a survey estimate from each area where the survey estimate has a distribution based on NEF-QVF, and apply the results in the previous section to provide a second-order approximation and its unbiased estimator for the conditional MSE of the EB. In our settings, it is assumed that the known parameters n_i's, which correspond to sample sizes in small-areas in normal cases, are bounded and the number of areas m is large.

3.1 Empirical Bayes estimator in NEF-QVF

Let $y_1, \ldots, y_m$ be mutually independent random variables where the conditional distribution of y_i given θ_i and the marginal distribution of θ_i belong to the the following natural exponential families:

$$\begin{aligned} y_i|\theta_i \sim & f(y_i|\theta_i) = \exp[n_i(\theta_i y_i - \psi(\theta_i)) + c(y_i, n_i)], \\ \theta_i|\nu, m_i \sim & \pi(\theta_i|\nu, m_i) = \exp[\nu(m_i\theta_i - \psi(\theta_i))]C(\nu, m_i), \end{aligned} \tag{13}$$

where n_i is a known scalar parameter and ν is an unknown scalar hyperparameter. Let $\boldsymbol{y} = (y_1, \ldots, y_m)^t$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^t$. The function $f(y_i|\theta_i)$ is the regular one-parameter exponential family and the function $\pi(\theta_i|\nu, m_i)$ is the conjugate prior distribution. Define ξ_i by

$$\xi_i = E[y_i|\theta_i] = \psi'(\theta_i),$$

which is the conditional expectation of y_i given θ_i. Assume that $\psi''(\theta_i) = Q(\xi_i)$, namely,

$$Var(y_i|\theta_i) = \frac{\psi''(\theta_i)}{n_i} = \frac{Q(\xi_i)}{n_i},$$

where $Q(x) = v_0 + v_1 x + v_2 x^2$ for known constants v_0, v_1 and v_2 which are not simultaneously zero. This means that the conditional variance $Var(y_i|\theta_i)$ is quadratic function of the conditional expectation $E[y_i|\theta_i]$. This is the natural exponential family with the quadratic variance function (NEF-QVF) introduced and investigated by Morris (1982, 1983). Similarly, the mean and variance of the prior distribution are given by

$$E[\xi_i|m_i, \nu] = m_i, \quad Var(\xi_i|m_i, \nu) = \frac{Q_i(m_i)}{\nu - v_2}. \tag{14}$$

In our settings, we consider the link given by

$$m_i = \psi'(\boldsymbol{x}_i^t \boldsymbol{\beta}), \quad i = 1, \ldots, m,$$

where $\boldsymbol{x}_i$ is a $p \times 1$ vector of explanatory variables and $\boldsymbol{\beta}$ is a $p \times 1$ unknown common vector of regression coefficients. Then, the unknown parameters $\boldsymbol{\eta}$ in the previous section correspond to $\boldsymbol{\eta}^t = (\boldsymbol{\beta}^t, \nu)$. The joint probability density (or mass) function of (y_i, θ_i) can be expressed as

$$f(y_i|\theta_i)\pi(\theta_i|\nu, m_i) = \pi(\theta_i|y_i, \nu) f_\pi(y_i|\nu, m_i),$$

where $\pi(\theta_i|y_i, \nu)$ is the conditional (or posterior) density function of θ_i given y_i, and $f_\pi(y_i|\nu, m_i)$ is the marginal density function of y_i. These density (or mass) functions are written as

$$\begin{aligned} \pi(\theta_i|y_i, \nu, m_i) =& \exp[(n_i + \nu)(\widehat{\xi}_i \theta_i - \psi(\theta_i))] C(n_i + \nu, \widehat{\xi}_i), \\ f_\pi(y_i|\nu, m_i) =& \frac{C(\nu, m_i)}{C(n_i + \nu, \widehat{\xi}_i)} \exp[c(y_i, n_i)], \end{aligned} \tag{15}$$

where $\widehat{\xi}_i$ is the posterior expectation of ξ_i, namely, $\widehat{\xi}_i = E[\xi_i|y_i, \boldsymbol{\eta}]$, given by

$$\widehat{\xi}_i = \widehat{\xi}_i(y_i, \boldsymbol{\eta}) = \frac{n_i y_i + \nu m_i}{n_i + \nu}, \tag{16}$$

which corresponds to the Bayes estimator of ξ_i in the Bayesian context when ν and m_i are known. As shown in Ghosh and Maiti (2004),

$$\begin{aligned} E[y_i] =& E[\psi'(\theta_i)] = m_i, \\ Var(y_i) =& Var(E[y_i|\theta_i]) + E[Var(y_i|\theta_i)] = Var(\xi_i) + E[Q_i(\xi_i)/n_i] = Q_i(m_i)\phi_i, \\ Cov(y_i, \xi_i) =& E[Cov(y_i, \xi_i)|\theta_i] + Cov(E[y_i|\theta_i], \xi_i) = Q_i(m_i)/(\nu - v_2), \end{aligned}$$

for $\phi_i = (1 + \nu/n_i)/(\nu - v_2)$. Using these observations, Ghosh and Maiti (2004) showed that the Bayes estimator $\widehat{\xi}_i$ given in (18) is the best linear unbiased predictor (BLUP) of ξ_i in terms of MSE.

Since the hyperparameters $\boldsymbol{\eta}$ are unknown, we need to estimate them from the joint marginal distribution of $\boldsymbol{y}$. For the purpose, Ghosh and Maiti (2004) suggested the estimating equations given in Godambe and Thompson (1989). Let $\boldsymbol{g}_i = (g_{1i}, g_{2i})^t$ for $g_{1i} = y_i - m_i$ and $g_{2i} = (y_i - m_i)^2 - \phi_i Q_i(m_i)$. Let

$$\begin{aligned} \boldsymbol{D}_i^t =& Q_i(m_i) \begin{pmatrix} \boldsymbol{x}_i & Q_i'(m_i)\phi_i \boldsymbol{x}_i \\ 0 & -(1 + v_2/n_i)(\nu - v_2)^{-2} \end{pmatrix}, \\ \boldsymbol{\Sigma}_i =& \mathbf{Cov}\,(\boldsymbol{g}_i) = \begin{pmatrix} \mu_{2i} & \mu_{3i} \\ \mu_{3i} & \mu_{4i} - \mu_{2i}^2 \end{pmatrix}, \end{aligned}$$

and $|\boldsymbol{\Sigma}_i| = \mu_{4i}\mu_{2i} - \mu_{2i}^3 - \mu_{3i}^2$, where $\mu_{ri} = E[(y_i - m_i)^r]$, $r = 1, 2, \ldots$, and exact expressions of μ_{2i}, μ_{3i} and μ_{4i} are given below. Then, Ghosh and Maiti (2004) derived the estimating equations given by $\sum_{i=1}^m \boldsymbol{D}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{g}_i = \boldsymbol{0}$, which are written as

$$\begin{aligned} &\sum_{i=1}^m \frac{1}{|\boldsymbol{\Sigma}_i|} \Big[\{\mu_{4i} - \mu_{2i}^2 - \mu_{3i}\phi_i Q_i'(m_i)\} g_{1i} + \{\mu_{2i}\phi_i Q_i'(m_i) - \mu_{3i}\} g_{2i} \Big] Q_i(m_i)\boldsymbol{x}_i = \boldsymbol{0}, \\ &\sum_{i=1}^m \frac{1}{|\boldsymbol{\Sigma}_i|} \{\mu_{2i} g_{2i} - \mu_{3i} g_{1i}\} Q_i(m_i)(1 + v_2/n_i)(\nu - v_2)^{-2} = 0. \end{aligned} \tag{17}$$

The equations can only be solved numerically. To accomplish this, we use the `optim` function in 'R' to solve the estimating equations by minimizing the sums of squares of the estimating functions, as noted in Ghosh and Maiti (2004). This approach may cause the problem in the presence of multiple roots, but fortunately we did not encounter this situation in our example. The exact moments $\mu_{ri} = E[(y_i - m_i)^r]$, $r = 1, 2, 3, 4$, are obtain from Theorem 1 of Ghosh and Maiti (2004) as

$$\mu_{2i} = \frac{Q(m_i)(\nu/n_i + 1)}{\nu - v_2}, \quad \mu_{3i} = \frac{Q(m_i)Q'(m_i)(\nu/n_i + 1)(\nu/n_i + 2)}{(\nu - v_2)(\nu - 2v_2)},$$

and

$$\begin{aligned}\mu_{4i} =&(d_i + 1)(2d_i + 1)(3d_i + 1)E[(\xi_i - m_i)^4] + \frac{6}{n_i}Q_i'(m_i)(d_i + 1)(2d_i + 1)E[(\xi_i - m_i)^3] \\ &+ \frac{d_i + 1}{n_i^2}\big[7\{Q'(m_i)\}^2 + 2n_i(4d_i + 3)Q(m_i)\big]E[(\xi_i - m_i)^2] \\ &+ \frac{1}{n_i^3}Q(m_i)\big[n_i(2d_i + 3)Q(m_i) + \{Q'(m_i)\}^2\big],\end{aligned}$$

for $d_i = v_2/n_i$. The expression of the moments of ξ_i are obtained given in Kubokawa, *et al.* (2014) as $E[(\xi_i - m_i)^2] = Q(m_i)/(\nu - v_2)$, $E[(\xi_i - m_i)^3] = 2Q(m_i)Q'(m_i)/(\nu - v_2)(\nu - 2v_2)$ and

$$E\left[(\xi_i - m_i)^4\right] = \frac{3Q(m_i)\left[(\nu - v_2)Q(m_i) + 2\{Q'(m_i)\}^2\right]}{(\nu - v_2)(\nu - 2v_2)(\nu - 3v_2)}.$$

Using these expressions, we obtain the estimator $\widehat{\boldsymbol{\eta}}^t = (\widehat{\boldsymbol{\beta}}^t, \widehat{\nu})$. Letting $\widehat{m}_i = \psi'(\boldsymbol{x}_i^t\widehat{\boldsymbol{\beta}})$ and substituting $\widehat{m}_i$ and $\hat{\nu}$ into (16), we finally get the empirical Bayes estimator of ξ_i, given by

$$\widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}}) = \frac{n_i y_i + \widehat{\nu}\widehat{m}_i}{n_i + \widehat{\nu}}. \tag{18}$$

The EB estimator is often used as a predictor in small area estimation and its uncertainty is of great importance. Our interest is in evaluation of the conditional MSE of $\widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}})$, which is investigated in the next subsection.

3.2 Evaluation of the conditional MSE

We begin by giving a stochastic expansion and conditional moments of $\widehat{\boldsymbol{\eta}}$ which is the solution of the estimating equations (17). We use the notations given by

$$\begin{aligned}\boldsymbol{s}_m =& \sum_{i=1}^m \boldsymbol{D}_i^t \boldsymbol{\Sigma}_i^{-1}\boldsymbol{g}_i, \\ \boldsymbol{U}(\boldsymbol{\eta}) =& \mathbf{Cov}\,(\boldsymbol{s}_m) = \sum_{i=1}^m \boldsymbol{D}_i^t \boldsymbol{\Sigma}_i^{-1}\boldsymbol{D}_i, \\ \boldsymbol{b}(\boldsymbol{\eta}) =& \boldsymbol{U}(\boldsymbol{\eta})^{-1}\Big(\boldsymbol{a}_1(\boldsymbol{\eta}) + \frac{1}{2}\boldsymbol{a}_2(\boldsymbol{\eta})\Big),\end{aligned}$$

where the detailed forms of $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are given in the Appendix. It is noted that $\boldsymbol{s}_m = O_p(m)$ and $\boldsymbol{U}(\boldsymbol{\eta}) = O(m)$. The following lemma is useful for evaluating the conditional MSE, where the proof is given in the Appendix.

Lemma 1. *Let $\widehat{\boldsymbol{\eta}}$ be the solution of estimating equations in* (17)*. Then for $i = 1, \ldots, m$,*

$$\begin{aligned}
&(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})|y_i = \boldsymbol{U}(\boldsymbol{\eta})^{-1}\boldsymbol{s}_m + o_p(m^{-1/2}),\\
&E[(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t|y_i] = \boldsymbol{U}(\boldsymbol{\eta})^{-1} + o_p(m^{-1}),\\
&E[\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}|y_i] = \boldsymbol{b}(\boldsymbol{\eta}) + o_p(m^{-1}).
\end{aligned} \tag{19}$$

Lemma 1 means that the second-order approximations of the conditional moments $E[(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t|y_i]$ and $E[\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}|y_i]$ do not depend on y_i, that is , they are equal to the unconditional moments given in Ghosh and Maiti (2004). Lemma 1 shows that the estimator $\widehat{\boldsymbol{\eta}}$ defined as the solution of (17) satisfies conditions (A.1) and (A.2).

We now derive analytical expressions of $T_{1i}(y_i, \boldsymbol{\eta})$ and $T_{2i}(y_i, \boldsymbol{\eta})$ in Theorem 1. Let

$$T_{1i}(y_i, \boldsymbol{\eta}) = \frac{Q(\widehat{\xi}_i(y_i, \boldsymbol{\eta}))}{n_i + \nu - v_2}, \qquad i = 1, \ldots, m,$$

which is $O_p(1)$. Let

$$T_{2i}(y_i, \boldsymbol{\eta}) = \mathrm{tr}\,\left[\boldsymbol{P}_i(y_i, \boldsymbol{\eta})\boldsymbol{U}(\boldsymbol{\eta})^{-1}\right],$$

which is $O_p(m^{-1})$, where

$$\boldsymbol{P}_i(y_i, \boldsymbol{\eta}) = (n_i + \nu)^{-2}\begin{pmatrix} \nu^2 Q(m_i)^2 \boldsymbol{x}_i\boldsymbol{x}_i^t & -n_i\nu(n_i + \nu)^{-1}Q(m_i)g_{1i}\boldsymbol{x}_i \\ -n_i\nu(n_i + \nu)^{-1}Q(m_i)g_{1i}\boldsymbol{x}_i^t & n_i^2(n_i + \nu)^{-2}g_{1i}^2 \end{pmatrix}.$$

Theorem 4. *The conditional mean squared error of $\widehat{\xi}(y_i, \widehat{\boldsymbol{\eta}})$ can be approximated up to $O_p(m^{-1})$ as*

$$\mathrm{cMSE}_i(\boldsymbol{\eta}, \widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}})|y_i) = T_{1i}(y_i, \boldsymbol{\eta}) + T_{2i}(y_i, \boldsymbol{\eta}) + o_p(m^{-1}). \tag{20}$$

Proof. From Theorem 1, it is sufficient to calculate T_{1i} and T_{2i}. It is easy to see that

$$Var(\xi_i|y_i) = \frac{Q(\widehat{\xi}_i(y_i, \boldsymbol{\eta}))}{n_i + \nu - v_2}, \qquad i = 1, \ldots, m,$$

so that we have the expression $T_{1i}(y_i, \boldsymbol{\eta})$. For $T_{2i}(y_i, \boldsymbol{\eta})$, we have

$$\begin{aligned}
E\Big[\Big\{(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t\frac{\partial\widehat{\xi}_i(y_i, \boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\Big\}^2\Big|y_i\Big] &= \mathrm{tr}\,E\Big[\Big(\frac{\partial\widehat{\xi}_i}{\partial\boldsymbol{\eta}}\Big)\Big(\frac{\partial\widehat{\xi}_i}{\partial\boldsymbol{\eta}}\Big)^t(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t\Big|y_i\Big]\\
&= \mathrm{tr}\,\Big[\Big(\frac{\partial\widehat{\xi}_i}{\partial\boldsymbol{\eta}}\Big)\Big(\frac{\partial\widehat{\xi}_i}{\partial\boldsymbol{\eta}}\Big)^t E\left[(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t|y_i\right]\Big].
\end{aligned}$$

It is noted from (16) that

$$\frac{\partial\widehat{\xi}_i(y_i, \boldsymbol{\eta})}{\partial\boldsymbol{\eta}} = \begin{pmatrix} \nu(n_i + \nu)^{-1}Q(m_i)\boldsymbol{x}_i \\ -n_i(n_i + \nu)^{-2}g_{1i} \end{pmatrix}.$$

Then from Lemma 1, the last formula can be approximated as

$$\mathrm{tr}\,\left[\boldsymbol{P}_i(y_i, \boldsymbol{\eta})\boldsymbol{U}(\boldsymbol{\eta})^{-1}\right] + o_p(m^{-1}),$$

which completes the proof. □

Taking the expectation of cMSE_i with respect to y_i, one gets the unconditional MSE given in Theorem 1 of Ghosh and Maiti (2004) with $\delta_i = n_i^{-1}$. In fact,

$$\begin{aligned}
T_{1i}(\boldsymbol{\eta}) \equiv & E[T_{1i}(y_i,\boldsymbol{\eta})] = \frac{\nu}{(n_i+\nu)(\nu-v_2)}Q(m_i),\\
T_{2i}(\boldsymbol{\eta}) \equiv & E[T_{2i}(y_i,\boldsymbol{\eta})]\\
= & (n_i+\nu)^{-2}\mathrm{tr}\left[\begin{pmatrix} \nu^2 Q(m_i)^2\boldsymbol{x}_i\boldsymbol{x}_i^t & \mathbf{0}\\ \mathbf{0}^t & n_i(n_i+\nu)^{-1}Q(m_i)(\nu-v_2)^{-1}\end{pmatrix}\boldsymbol{U}(\boldsymbol{\eta})^{-1}\right].
\end{aligned}$$

Corollary 1. *The unconditional MSE of $\widehat{\xi}_i(y_i,\widehat{\boldsymbol{\eta}})$ is approximated as*

$$\mathrm{MSE}(\boldsymbol{\eta},\widehat{\xi}_i(y_i,\widehat{\boldsymbol{\eta}})) = T_{1i}(\boldsymbol{\eta}) + T_{2i}(\boldsymbol{\eta}) + o(m^{-1}). \tag{21}$$

It is interesting to investigate the difference between the approximations of the cMSE and the MSE. When the underlying distribution of y_i is a normal distribution, we have $Q(x)=1$, or $v_0=1$ and $v_1=v_2=0$, so that $T_{1i}(y_i,\boldsymbol{\eta}) = 1/(n_i+\nu) = T_{1i}(\boldsymbol{\eta})$, namely the leading term in the cMSE is identical to that in the MSE. Thus, the difference between the cMSE and the MSE appears in the second-order term with $O_p(m^{-1})$. When v_1 or v_2 is not zero, however, the leading term $T_{1i}(y_i,\boldsymbol{\eta})$ in the cMSE is a function of y_i and it is not equal to the leading term $T_{1i}(\boldsymbol{\eta})$ in the MSE. Thus, for distributions far from the normality, the difference between the cMSE and the MSE is significant even when m is large. This tells us about the remark that one cannot replace the the conditional MSE given y_i with the corresponding unconditional MSE except for the normal distribution. Some examples including the Poisson and binomial distributions are given in Section 3.3.

We next derive an analytical form of a second-order unbiased estimator for the cMSE. For the purpose, we need to calculate T_{11i} and T_{12i} given in (8) and (9), respectively. Note that

$$\begin{aligned}
\boldsymbol{r}(y_i,\boldsymbol{\eta}) \equiv & \frac{\partial T_{1i}}{\partial \boldsymbol{\eta}} = \begin{pmatrix} \nu(n_i+\nu)^{-1}\lambda_i Q'(\widehat{\xi}_i)Q(m_i)\boldsymbol{x}_i \\ -\lambda_i^2 Q(\widehat{\xi}_i) - \lambda_i n_i(n_i+\nu)^{-2}Q'(\widehat{\xi}_i)g_{1i}\end{pmatrix},\\
\boldsymbol{R}(y_i,\boldsymbol{\eta}) \equiv & \frac{\partial^2 T_{1i}}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^t} = \begin{pmatrix} \boldsymbol{T}_{1i}^{11} & \boldsymbol{T}_{1i}^{12}\\ (\boldsymbol{T}_{1i}^{12})^t & T_{1i}^{22}\end{pmatrix},
\end{aligned}$$

where $\lambda_i = (n_i+\nu-v_2)^{-1}$, and

$$\begin{aligned}
\boldsymbol{T}_{1i}^{11} = & (n_i+\nu)^{-2}\nu\boldsymbol{x}_i\boldsymbol{x}_i^t\lambda_i Q(m_i)\left[2v_2\nu Q(m_i) + Q'(\widehat{\xi}_i)Q'(m_i)(n_i+\nu)\right],\\
\boldsymbol{T}_{1i}^{12} = & \frac{\partial^2 T_{1i}}{\partial\boldsymbol{\beta}\partial\nu} = Q(m_i)\lambda_i(n_i+\nu)^{-2}\left\{Q'(\widehat{\xi}_i)\left(n_i-\nu(n_i+\nu)\lambda_i\right) - 2v_2 n_i\nu g_{1i}(n_i+\nu)^{-1}\right\}\boldsymbol{x}_i,\\
T_{1i}^{22} = & \frac{\partial^2 T_{1i}}{\partial\nu^2} = 2\lambda_i^3 Q(\widehat{\xi}_i) + 2\lambda_i^2 n_i(n_i+\nu)^{-2}Q'(\widehat{\xi}_i)g_{1i}\\
& + 2\lambda_i n_i(n_i+\nu)^{-4}g_{1i}\left[(n_i+\nu)Q'(\widehat{\xi}_i) + n_i v_2 g_{1i}\right].
\end{aligned}$$

Using (19) in Lemma 1, we obtain the analytical expression of T_{11i} and T_{12i} as

$$\begin{aligned}
T_{11i}(y_i,\boldsymbol{\eta}) = & \boldsymbol{r}(y_i,\boldsymbol{\eta})^t\boldsymbol{b}(\boldsymbol{\eta}),\\
T_{12i}(y_i,\boldsymbol{\eta}) = & \frac{1}{2}\mathrm{tr}\left[\boldsymbol{R}(y_i,\boldsymbol{\eta})\boldsymbol{U}(\boldsymbol{\eta})^{-1}\right].
\end{aligned}$$

The estimator $\widehat{\text{cMSE}}_i$ given in (10) is expressed as

$$\widehat{\text{cMSE}}_i(\widehat{\xi}_i(y_i,\widehat{\boldsymbol{\eta}})) = T_{1i}(y_i,\widehat{\boldsymbol{\eta}}) + T_{2i}(y_i,\widehat{\boldsymbol{\eta}}) - \boldsymbol{r}(y_i,\widehat{\boldsymbol{\eta}})^t\boldsymbol{b}(\widehat{\boldsymbol{\eta}}) - \frac{1}{2}\text{tr}\left[\boldsymbol{R}(y_i,\widehat{\boldsymbol{\eta}})\boldsymbol{U}(\widehat{\boldsymbol{\eta}})^{-1}\right]. \tag{22}$$

Theorem 5. *The estimator* (22) *is a second-order unbiased estimator, namely,*

$$E[\widehat{\text{cMSE}}_i(\widehat{\xi}_i(y_i,\widehat{\boldsymbol{\eta}}))] = \text{cMSE}_i(\boldsymbol{\eta},\widehat{\xi}_i(y_i,\widehat{\boldsymbol{\eta}})|y_i) + o_p(m^{-1}).$$

It is noted that the results in Theorems 4 and 5 do **not** require the condition that $n_i\to\infty$, while the condition is assumed in Booth and Hobert (1998) for the generalized linear mixed model. Thus, the results in Theorems 4 and 5 are applicable in the context of small area estimation.

3.3 Some examples

We give some examples of the mixed models belonging to (13) and investigate the conditional MSE.

[1] **Fay-Herriot model**. The Fay-Herriot model is an area-level model often used in small area estimation, given by

$$y_i = \boldsymbol{x}_i^t\boldsymbol{\beta} + v_i + \varepsilon_i, \qquad i=1,\ldots,m,$$

where m is the number of small areas, and v_i's and ε_i's are mutually independently distributed random errors such that $v_i\sim\mathcal{N}(0,A)$ and $\varepsilon_i\sim\mathcal{N}(0,D_i)$. The notations in (13) correspond to $n_i = D_i^{-1}, v_0=1, v_1=v_2=0, \xi_i=\theta_i, \nu=A^{-1}$ and $\psi(\theta_i)=\theta_i^2/2$. In this case, the estimating equations in (17) reduce to

$$\begin{aligned}
&\sum_{i=1}^m (A+D_i)^{-1}\boldsymbol{x}_i y_i = \sum_{i=1}^m (A+D_i)^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^t\boldsymbol{\beta},\\
&\sum_{i=1}^m (A+D_i)^{-2}(y_i-\boldsymbol{x}_i^t\boldsymbol{\beta})^2 = \sum_{i=1}^m (A+D_i)^{-1},
\end{aligned}$$

which coincide with the likelihood equations for the maximum likelihood estimators of $\boldsymbol{\beta}$ and A. The terms $T_{1i}(y_i,\boldsymbol{\eta})$ and $T_{2i}(y_i,\boldsymbol{\eta})$ in approximation (20) of the cMSE are written as

$$\begin{aligned}
T_{1i}(y_i,\boldsymbol{\eta}) &= \frac{AD_i}{A+D_i}\\
T_{2i}(y_i,\boldsymbol{\eta}) &= \frac{D_i}{(A+D_i)^2}\boldsymbol{x}_i^t\left(\sum_{j=1}^m \frac{\boldsymbol{x}_j\boldsymbol{x}_j^t}{A+D_j}\right)^{-1}\boldsymbol{x}_j + \frac{D_i^2(y_i-\boldsymbol{x}_i^t\boldsymbol{\beta})^2}{(A+D_j)^4}\left(\sum_{j=1}^m\frac{1}{2(A+D_j)^2}\right)^{-1},
\end{aligned}$$

which were given in Datta *et al* (2011). In the Fay-Herriot model, $T_{1i}(y_i,\boldsymbol{\eta}) = AD_i/(A+D_i) = T_{1i}(\boldsymbol{\eta})$, namely, the leading terms in the cMSE and the MSE are identical, and the difference between the cMSE and MSE is small for large m.

[2] Poisson-gamma mixture model. Let $z_1, \ldots, z_m$ be mutually independent random variables having

$$z_i|\lambda_i \sim \text{Po}(n_i\lambda_i) \quad \text{and} \quad \lambda_i \sim \text{Ga}(\nu m_i, 1/\nu)$$

where $\lambda_1, \ldots, \lambda_m$ are mutually independent, $\text{Po}(\lambda)$ denotes the Poisson distribution with mean λ, and $\text{Ga}(a, b)$ denotes the gamma distribution with shape parameter a and scale parameter b. Let $y_i = z_i/n_i$ and $\log m_i = \boldsymbol{x}_i^t\boldsymbol{\beta}$ for $i = 1, \ldots, m$. Then, the notations in (13) correspond to $v_1 = 1$, $v_0 = v_2 = 0$, $\xi_i = \lambda_i = \exp(\theta_i)$, and $\psi(\theta_i) = \exp(\theta_i)$. The posterior distribution of λ_i is $\text{Ga}(\nu m_i + n_i y_i, (n_i + \nu)^{-1})$ or $\text{Ga}((n_i + \nu)\widehat{\xi}_i, (n_i + \nu)^{-1})$. Then we have

$$T_{1i}(y_i, \boldsymbol{\eta}) = \frac{\widehat{\xi}(y_i, \boldsymbol{\eta})}{n_i + \nu} = \frac{n_i y_i + \nu m_i}{(n_i + \nu)^2},$$

which increases in y_i. Thus, the difference between the unconditional and conditional MSE gets large as y_i gets large. When a large value of y_i is observed, it should be remarked that the conditional prediction error of the empirical Bayes estimator given y_i is larger than the unconditional (or integrated) prediction error. Hence, it is meaningful to provide to practitioners the information on the conditional MSE as well as the unconditional MSE.

[3] Binomial-beta mixture model. Let $z_1, \ldots, z_m$ be mutually independent random variables having

$$z_i|p_i \sim \text{Bin}(n_i, p_i) \quad \text{and} \quad p_i \sim \text{Beta}(\nu m_i, \nu(1 - m_i)),$$

where $p_1, \ldots, p_m$ are mutually independent, $\text{Bin}(n, p)$ denotes the binomial distribution and $\text{Beta}(a, b)$ denotes the beta distribution. Let $y_i = z_i/n_i$ and $m_i = \exp(\boldsymbol{x}_i^t\boldsymbol{\beta})/(1 + \exp(\boldsymbol{x}_i^t\boldsymbol{\beta}))$ for $i = 1, \ldots, m$. Then the notations in (13) correspond to $v_0 = 0$, $v_1 = 1$ and $v_2 = -1$, $\xi_i = p_i = \exp(\theta_i)/(1 + \exp(\theta_i))$ and $\psi(\theta_i) = \log(1 + \exp(\theta_i))$. The posterior distribution of p_i is $\text{Beta}(\nu m_i + n_i y_i, n_i(1 - y_i) + \nu(1 - m_i))$ or $\text{Beta}((n_i + \nu)\widehat{\xi}_i, (n_i + \nu)(1 - \widehat{\xi}_i))$, so that $T_{1i}(y_i, \boldsymbol{\eta})$ is written as

$$T_{1i}(y_i, \boldsymbol{\eta}) = \frac{\widehat{\xi}_i(y_i, \boldsymbol{\eta})(1 - \widehat{\xi}_i(y_i, \boldsymbol{\eta}))}{n_i + \nu + 1},$$

which is a quadratic and concave function of y_i. Since $0 < \widehat{\xi}(y_i, \boldsymbol{\eta}) < 1$, $T_{1i}(y_i, \boldsymbol{\eta})$ is always positive and attains the maximum when $\widehat{\xi}_i = 1/2$ or $y_i = (n_i + \nu)/2n_i - \nu m_i/n_i$, and $T_{1i}(y_i, \boldsymbol{\eta}) = 0$ when $\widehat{\xi}_i = 0$ or 1. Thus, the value of $T_{1i}(y_i, \boldsymbol{\eta})$ is relatively small when y_i is close to 0 or 1. When y_i is around 1/2, the value of $T_{1i}(y_i, \boldsymbol{\eta})$ tends to be larger. When a value around 1/2 is observed for y_i, it should be remarked that the conditional prediction error of the EB given y_i is larger than the unconditional (or integrated) prediction error.

4 Numerical and Empirical Studies

We here give some comparisons of the conditional and unconditional MSEs and investigate finite sample performances of the second-order unbiased estimator of the cMSE. We also apply the suggested procedures to real mortality data.

4.1 Comparison of the conditional and unconditonal MSEs

It is interesting to investigate how different the conditional MSE is from the unconditional MSE. The major difference between them appears in the leading terms, namely the terms with order $O_p(1)$ in the cMSE and MSE. The ratio of the leading term of the cMSE to that of the MSE is defined by

$$\mathrm{Ratio}_1 = T_{1i}(y_i, \boldsymbol{\eta})/E[T_{1i}(y_i, \boldsymbol{\eta})],$$

which is a function of y_i and $\boldsymbol{\eta}$. We consider the case that $m = 10$, $\nu = 1$, $\boldsymbol{x}_i^t\boldsymbol{\beta} = \mu = 0$ and $n_i = 10$ for $i = 1, \ldots, m$. Then, the curves of the functions Ratio_1 are illustrated Figure 1 for the three mixed models: the Fay-Herriot, Poisson-gamma and binomial-beta models. As mentioned before, in the Fay-Herriot (or normal-normal mixture) model, $\mathrm{Ratio}_1 = 1$ since $T_{1i}(y_i, \boldsymbol{\eta}) = E\left[T_{1i}(y_i, \boldsymbol{\eta})\right]$. For the Poisson-gamma and binomial-beta mixture models, Figure 1 tells us about the interesting features of their leading terms in the cMSE, namely, the ratio is an increasing function of y_i for the Poisson-gamma mixture model, and a concave and quadratic function of y_i for the binomial-beta mixture model.

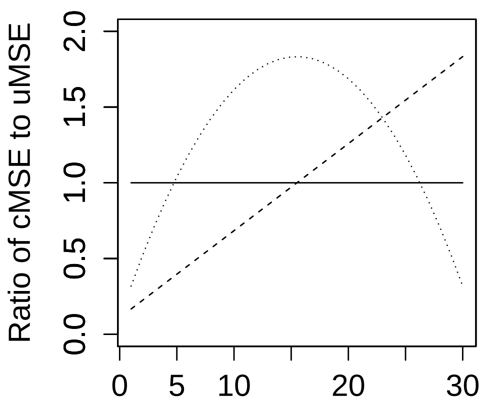

Figure 1: *Figures of* Ratio_1 *for the Three Mixed Models (The solid, dashed and dotted lines correspond to the Fay-Herriot, Poisson-gamma mixture and binomial-beta mixture models, respectively.)*

We next investigate the corresponding ratios based on the second-order approximations of the cMSE and MSE. Let us define Ratio_2 by

$$\mathrm{Ratio}_2 = \{T_{1i}(y_i, \boldsymbol{\eta}) + T_{2i}(y_i, \boldsymbol{\eta})\}/E[T_{1i}(y_i, \boldsymbol{\eta}) + T_{2i}(y_i, \boldsymbol{\eta})],$$

where $T_{1i}(y_i, \boldsymbol{\eta}) + T_{2i}(y_i, \boldsymbol{\eta})$ and $E[T_{1i}(y_i, \boldsymbol{\eta}) + T_{2i}(y_i, \boldsymbol{\eta})]$ are given in (20) and (21), respectively. Since the second-order terms depend on m, we treat the three cases of $m = 10$, 15 and 20 for $\boldsymbol{x}_i'\boldsymbol{\beta} = \mu$ and $n_1 = \cdots = n_m = 5$. The performaces of Ratio_2 are illustrated in Figure 2 for the three mixed models, where the values of (μ, ν) are $(0, 1)$ for the Fay-Herriot model, $(\exp(2), 0)$

for the Poisson-gamma mixture model, and $(\exp(1.5)/(1+\exp(1.5)), 0)$ for the binomial-beta mixture models. Figure 2 demonstrates that the second-order terms for the three mixed models do not contribute so much to Ratio_2 or the conditional MSE.

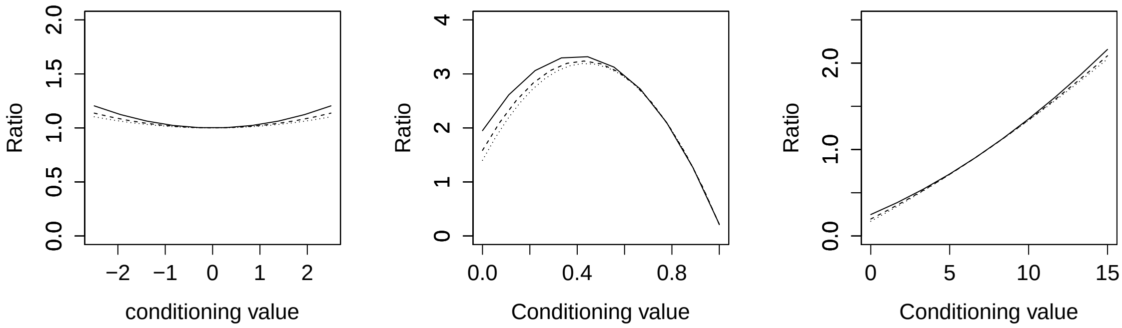

Figure 2: *Figures of* Ratio_2 *for the Fay-Herriot Model (Left), the Binomial-beta Mixture Model (Center) and the Poisson-gamma Mixture Model (Right) (The solid, dashed and dotted lines correspond to the cases of* $m = 10$, 15 *and* 20, *respectively. The conditioning value denotes* y_i.*)*

4.2 Finite performances of the estimator of cMSE

We investigate finite performances of the second-order unbiased estimator for the conditional MSE by simulation. The mixed models we examin are the Poisson-gamma mixture and binomial-beta mixture models where the simple case of $\boldsymbol{x}_i'\boldsymbol{\beta} = 0$ without covariates is treated with $m = 25$, $n_i = 10$ and $\nu = 15$.

In the experiment of simulation, let $i = 1$ be the index for the area of interest, namely the value of y_1 is conditioned. As seen from the discussion given in Section 4.1, the performances of the conditional MSE depend on the value of y_1. In this simulation, we consider the α-quantile point, denoted by $y_{1(\alpha)}$, of the distribution of y_1 and select the five quantiles $y_{1(\alpha)}$ for $\alpha = 0.05$, 0.25, 0.5, 0.75 and 0.95. For the Poisson-gamma mixture model, the marginal distribution of y_1 is the negative binomial distribution $NB(\nu m_1, \nu/(n_1+\nu))$, and we can obtain the five quantiles $y_{1(\alpha)}$ from the marginal distribution. For the binomial-beta mixture model, the marginal distribution of y_1 is not given as a typical distribution. Thus, we need to calculate numerically α-quantile values of y_1.

The true values of cMSE can be provided based on the simulation with $R = 10,000$ replications. For $r = 1, \ldots, R$, we generate random variables $y_i^{(r)}$ and $\theta_i^{(r)}$, $i = 2, \ldots, m$, which are distributed as $y_i^{(r)}|(\theta_i^{(r)}, \mu, \nu) \sim f(y_i|\theta_i^{(r)}, \mu, \nu)$ and $\theta_i^{(r)}|(\mu, \nu) \sim \pi(\theta_i|\mu, \nu)$. In the r-th replication, from the sample $\{y_{1(\alpha)}, y_2^{(r)}, \ldots, y_m^{(r)}\}$, we calculate the values of $\widehat{\xi}_1(y_{1(\alpha)}, \widehat{\boldsymbol{\eta}})^{(r)}$ and $\widehat{\xi}_1(y_{1(\alpha)}, \boldsymbol{\eta})^{(r)}$. Then, the true value of the cMSE of $\widehat{\xi}_1(y_{1(\alpha)}, \widehat{\boldsymbol{\eta}})$ can be numerically calculated as

$$\mathrm{cMSE}_1 = T_{11}(y_{1(\alpha)}, \boldsymbol{\eta}) + \frac{1}{R}\sum_{r=1}^{R}\Big\{\widehat{\xi}_1(y_{1(\alpha)}, \widehat{\boldsymbol{\eta}})^{(r)} - \widehat{\xi}_1(y_{1(\alpha)}, \boldsymbol{\eta})^{(r)}\Big\}^2.$$

Through the same manner as described above, we generate another simulated sample with size $T = 2,000$ and calculate the cMSE estimate $\widehat{\text{cMSE}}_1$ from (22). Then, we can obtain the relative bias (RB) and coefficients of variation (CV) for the cMSE estimator, which are defined by

$$\begin{aligned}
\text{RB} =& \frac{T^{-1}\sum_{t=1}^{T}\widehat{\text{cMSE}}_1^{(t)} - \text{cMSE}_1}{\text{cMSE}_1}, \\
\text{CV} =& \Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\widehat{\text{cMSE}}_1^{(t)} - \text{cMSE}_1\Big)^2\Big]^{1/2}\Big/\text{cMSE}_1,
\end{aligned}$$

where $\widehat{\text{cMSE}}_1^{(t)}$ denotes the cMSE estimate in the t-th replication for $t = 1, \ldots, T$.

For $\alpha = 0.05$, 0.25, 0.50, 0.75 and 0.95, the values of $y_{1(\alpha)}$, cMSE_1, $E[\widehat{\text{cMSE}}_1]$, RB and CV are reported in Table 1 for the two mixed models, where the values of cMSE_1 and $E[\widehat{\text{cMSE}}_1]$ are multiplied by 100. Table 1 demonstrates that the estimator $\widehat{\text{cMSE}}_1$ of the conditional MSE performs well for various values of $y_{1(\alpha)}$in both models. The true value of cMSE_i has a general treand of increase in $y_{1(\alpha)}$ for the Poisson-gamma mixture model, and this conincides with the analytical property discussed in Section 4.1. For the binomial-beta mixture model, the true values of cMSE_i are about the same and do have a feature of concave explained in Section 4.1. The values of RB and CV show that the analytical second-order unbiased estimator given in (22) is not bad as an etimator of cMSE_i.

Table 1: *Values of* cMSE_1, $E[\widehat{\text{cMSE}}_1]$, *Relative Bias* (RB) *and Coefficient of Variation* (CV) *of the cMSE Estimator for the Five Conditioning Values in the Poisson-gamma and Binomial-beta Mixture Models*

	α	$y_{1(\alpha)}$	cMSE_1	$E[\widehat{\text{cMSE}}_1]$	RB	CV
	0.05	0.40	4.10	4.53	0.10	0.75
	0.25	0.70	3.80	3.88	0.02	0.57
Poisson-gamma	0.50	1.00	4.24	4.14	-0.02	0.70
	0.75	1.30	4.90	4.67	-0.05	0.71
	0.95	1.70	6.16	6.58	0.07	0.60
	0.05	0.10	1.18	1.06	-0.10	0.28
	0.25	0.30	1.07	1.10	0.03	0.49
binomial-beta	0.50	0.40	1.03	1.12	0.09	0.58
	0.75	0.50	1.03	1.09	0.06	0.62
	0.95	0.70	1.06	1.05	-0.01	0.51

4.3 Empirical examples

We now apply the suggested procedures to the two data sets: the Stomach Cancer Mortality Data and the Infant Mortality Data Before World War II, both of which are data from prefectures in Japan.

Example 1 (Mortality rates estimates in the Poisson-gamma mixture model)**.** We begin by analyzing the Stomach Cancer Mortality Data in Japan. The data set consists of the observed number of mortality z_i and its expected number n_i of stomach cancer for women who lived in the i-th city or town in Saitama prefecture, Japan, for five years from 1995 to 1999. Such area-level data (z_i, n_i), $i = 1, \ldots, m$, are available for $m = 92$ cities and towns, and the total number of mortality in the whole region is $L = 3953$. The expected numbers are adjusted by age on the basis of the population so that $L = \sum_{i=1}^m z_i = \sum_{i=1}^m n_i$.

For $z_1, \ldots, z_K$, we use the Poisson-gamma mixture model discussed in Section 3.3, namely $z_i|\lambda_i \sim \mathrm{Po}(n_i\lambda_i)$ and $\lambda_i \sim \mathrm{Ga}(\nu m_i, 1/\nu)$. Since data of mortality rate of stomach cancer for men are also available, we can use them as a covariate. Let x_i be a log-transformed mortality rate for men for i-th area. Then, we treat the regression model $\log m_i = \beta_0 + x_i\beta_1$ for $i = 1, \ldots, m$. The unknown parameters $\boldsymbol{\eta}^t = (\beta_0, \beta_1, \nu)^t$ are estimated as the roots of the estimating equations in (17). Their estimates are $\beta_0 = -7.77 \times 10^{-3}$, $\beta_1 = 0.157$ and $\nu = 158$.

To illustrate the difference between cMSE and MSE, we use the percentage relative difference (RD) defined by

$$\mathrm{RD}_i = 100 \times (\widehat{\mathrm{cMSE}}_i - \widehat{\mathrm{MSE}}_i)/\widehat{\mathrm{MSE}}_i.$$

When RD_i is positive, $\widehat{\mathrm{cMSE}}_i$ is larger than $\widehat{\mathrm{MSE}}_i$, or MSE under-estimates the prediction error. When $\mathrm{RD}_i > 100$, $\widehat{\mathrm{cMSE}}_i$ is larger than twice the $\widehat{\mathrm{MSE}}_i$, and we should note the prediction error of the EB for given data. In Figure 3, the plots of the values $(\widehat{\mathrm{MSE}}_i, \widehat{\mathrm{cMSE}}_i)$ multiplied by $1,000$ and the values of (y_i, RD_i) for $i = 1, \ldots, m$ are given in the left and right figures, respectively, where $y_i = z_i/n_i$ is the standard mortality rate (SMR). From Figure 3, it is revealed that the values of $\widehat{\mathrm{cMSE}}_i$ are larger than those of $\widehat{\mathrm{MSE}}_i$ for many areas, and that the relative differences RD_i have great variability, which comes from non-normality of distribution as discussed in Section 4.1. For areas in which RD_i are large, we should note that the unconditional MSEs, which has been used conventionally in small area estimation, seem to under-estimate the prediction errors of the EB estimates for given observations of the areas.

Table 2 reports the values of n_i, y_i, EB_i, $\widehat{\mathrm{cMSE}}_i$, $\widehat{\mathrm{MSE}}_i$ and RD_i for ten selected municipalities in Saitama prefecture, where the values of $\widehat{\mathrm{MSE}}_i$ and $\widehat{\mathrm{cMSE}}_i$ are multiplied by $1,000$. It is noted that Minamikawara has the maximum RD value and Ageo has the minimum RD value in our result. The values of RD tell us about important information that the given empirical Bayes estimate has a larger prediction error than the usual unconditional MSE. For instance, Kamiizumi shows that SMR $y_i = 3.285$ is much shrunken to 1.050 by EB since $n_i = 1.5$ is very small. The estimate of the conditional MSE is 1.653, while that of the unconditional MSE is 0.767. The resulting RD is 116, quite high, which suggests that the unconditional MSE under-estimates the prediction error. Thus, we suggest to provide estimates of cMSE as well as estimates of MSE.

Example 2 (Infant mortality rates estimates in the binomial-beta mixture model)**.** We next handle the historical data of the Infant Mortality Data Before World War II. The data set

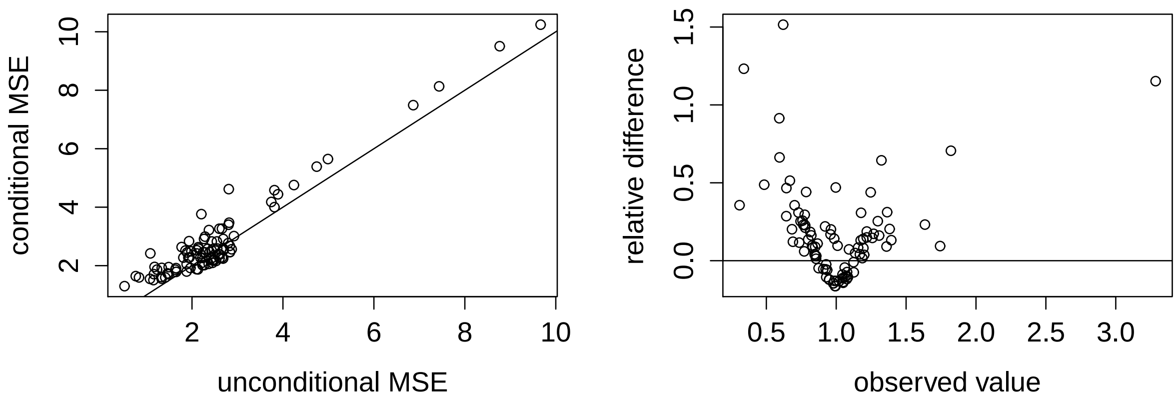

Figure 3: Plots of $(\widehat{\mathrm{MSE}}_i, \widehat{\mathrm{cMSE}}_i)$ (left) and Plots of (y_i, RD_i) (right) for Stomach Cancer Mortality Data

Table 2: *Values of* n_i*, SMR* y_i*,* EB_i*,* $\widehat{\mathrm{cMSE}}_i$*,* $\widehat{\mathrm{MSE}}_i$ *and* RD_i *for Selected Areas in Saitama Prefecture*

Area	n_i	y_i	EB_i	$\widehat{\mathrm{cMSE}}_i$	$\widehat{\mathrm{MSE}}_i$	RD_i
Ageo	110.0	0.991	0.994	2.282	2.741	-16
Asaka	52.5	1.124	1.005	2.839	2.918	-2
Hannou	58.2	0.979	0.989	2.221	2.634	-15
Kamiizumi	1.5	3.285	1.050	1.653	0.767	116
Kamikawa	10.1	0.594	0.993	1.966	1.185	66
Kumagaya	102.8	1.324	1.138	4.666	2.813	64
Minamikawara	3.2	0.620	0.979	1.314	0.535	146
Okano	11.8	0.339	0.954	2.431	1.098	121
Shiraoka	26.2	0.764	0.987	2.639	2.145	23
Yashio	37.5	0.828	0.981	2.583	2.353	10

consists of the observed number of infant mortality z_i and the number of birth n_i in the i-th city or town in Ishikawa prefecture, Japan, before World War II. Such area-level data are available for $m = 211$ cities, towns and villages, and the total number of infant mortality in the whole region is $L = 4252$.

It is noted that the infant mortality rates $y_i = z_i/n_i$ before World War II are not small and distributed around 0.2. Thus, we here apply the data to the binomial-beta model rather than the Poisson-gamma model. For $z_1, \ldots, z_K$, $z_i|p_i$ and p_i have the distributions $z_i|p_i \sim \mathrm{Bin}(n_i, p_i)$ and $p_i \sim \mathrm{Beta}(\nu m_i, \nu(1 - m_i))$, where $m_i = \exp(\beta)/(1 + \exp(\beta))$ for $i = 1, \ldots, m$, since we do not have any covariates. Thus, the unknown parameters are $\boldsymbol{\eta}^t = (\beta, \nu)^t$ and their estimates

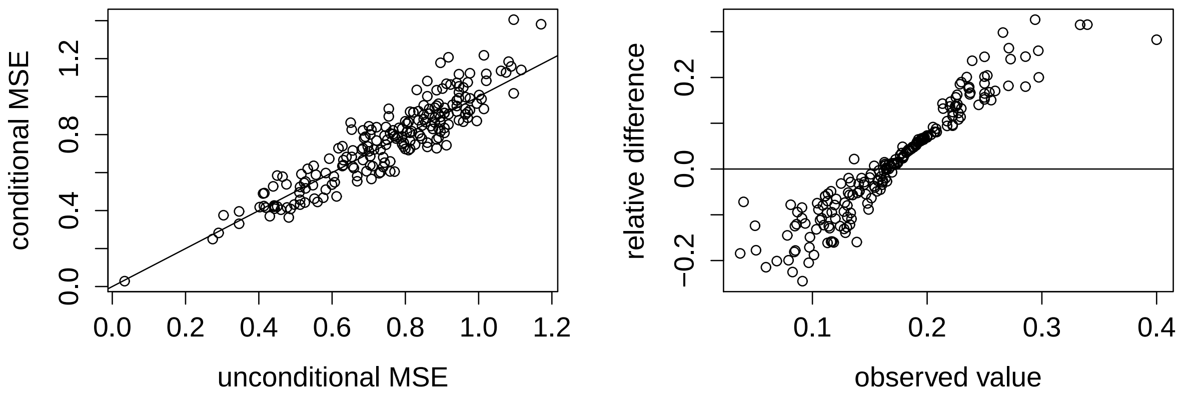

Figure 4: Plots of $(\widehat{\mathrm{MSE}}_i, \widehat{\mathrm{cMSE}}_i)$ (left) and Plots of (y_i, RD_i) (right) for Infant Mortality Data

are $\beta = -1.57$, namely $m_i = 0.171$, and $\nu = 102$.

The plots of the values $(\widehat{\mathrm{MSE}}_i, \widehat{\mathrm{cMSE}}_i)$ multiplied by $1,000$ and the values of (y_i, RD_i) for $i = 1, \ldots, m$ are given in the left and right figures of Figure 4, respectively. Figure 4 suggests that the values of the relative difference RD gets larger as y_i is larger. This is because the leading $O_p(1)$ term is an increasing function of y_i for fixed n_i since y_i is between 0 and 0.5, as investigated in Section 4.1. Table 3 reports the values of n_i, y_i, EB_i, $\widehat{\mathrm{cMSE}}_i$, $\widehat{\mathrm{MSE}}_i$ and RD_i for fifteen selected municipalities in Ishikawa prefecture, where the values of $\widehat{\mathrm{MSE}}_i$ and $\widehat{\mathrm{cMSE}}_i$ are multiplied by $1,000$. It is noted that Area 175 has the maximum RD value and Area 46 has the minimum RD value in our result. For Area 176, the observed mortality rate $y_i = 0.400$ is much shrunken to $\mathrm{EB}_i = 0.216$ by the empirical Bayes estimator since the number of birth is quite small as given by $n_i = 25$. The unconditional MSE is estimated by 1.123, but the relative difference is $\mathrm{RD}_i = 27$, and the estimate of cMSE is 1.436, which is higher than the MSE estimate. This suggests that it should be good to provide estimates of cMSE as well as estimates of MSE.

5 Concluding Remarks

In this paper, we have derived the second-order approximation of the conditional MSE of the empirical Bayes estimator and its second-order unbiased estimator in the general mixed models. Those results have been applied to the mixed models based on NEF-QVF, and the second-order evaluations of the cMSE have been provided in analytical and closed forms without assuming that the sample size n_i goes to infinity. It has been shown that the difference between the cMSE and the MSE is small for the normal distribution, while it is significant for the Poisson-gamma and the binomial-beta mixture models. We have also clarified how different the cMSE is from the MSE by comparing the leading terms in the cMSE and MSE. Through the empirical studies, the importance of cMSE has been illustrated for the Poisson-gamma and the binomial-beta mixture models, and we suggest to provide estimates of cMSE as well as estimates of

Table 3: *Values of* n_i, y_i, EB_i, $\widehat{\mathrm{cMSE}}_i$, $\widehat{\mathrm{MSE}}_i$ *and* RD_i *for Selected Areas in Ishikawa Prefecture*

Area	n_i	y_i	EB_i	$\widehat{\mathrm{cMSE}}_i$	$\widehat{\mathrm{MSE}}_i$	RD_i
1	4146	0.139	0.139	0.028	0.034	-15
19	56	0.250	0.199	1.060	0.916	15
23	55	0.164	0.168	0.928	0.921	0
46	197	0.091	0.119	0.366	0.485	-24
71	84	0.060	0.121	0.610	0.780	-21
79	87	0.069	0.124	0.611	0.768	-20
86	101	0.079	0.125	0.571	0.715	-20
96	194	0.119	0.137	0.411	0.490	-16
98	208	0.250	0.224	0.584	0.467	24
112	94	0.160	0.166	0.729	0.740	-1
158	173	0.185	0.180	0.551	0.527	4
162	57	0.333	0.229	1.199	0.910	31
175	119	0.294	0.237	0.874	0.657	33
176	25	0.400	0.216	1.436	1.123	27
179	245	0.229	0.212	0.496	0.417	18

MSE.

It should be beneficial that the second-order unbiased estimator of cMSE can be provided in a closed form in the mixed models based on NEF-QVF. As discussed in Booth and Hobert (1998), we cannot derive a second-order unbiased estimator with a closed form in the framework of the generalized linear mixed model, and we need to resort to numerical methods like bootstrap.

As an estimator of the hyperparameters, in this paper, we have used the solution of the estimating equations given in Ghosh and Maiti (2004). Their procedure allows us to express the asymptotic bias and variance in closed forms like Lemma 1. Another procedure is the maximum likelihood estimator (MLE) of the marginal likelihood. For the Poisson-gamma mixture model, the marginal distribution of y_i (marginal likelihood) is the negative binomial distribution given by

$$f(y_i|\boldsymbol{\eta}) = \frac{\Gamma(n_i y_i + \nu m_i)}{\Gamma(n_i y_i + 1)\Gamma(\nu m_i)} \left(\frac{n_i}{n_i + \nu}\right)^{n_i y_i} \left(\frac{\nu}{n_i + \nu}\right)^{\nu m_i},$$

where $\Gamma(\cdot)$ denotes a gamma function. For the binomial-beta mixture model, the marginal likelihood is proportional to

$$L(\boldsymbol{\eta}) \propto \prod_{i=1}^{m} \frac{B(\nu m_i + n_i y_i, n_i(1 - y_i) + \nu(1 - m_i))}{B(\nu m_i, \nu(1 - m_i))},$$

where $B(\cdot)$ denotes a beta function. Then, we can obtain MLE of the parameters as a maximizer of the marginal likelihood. However, it is difficult to derive the asymptotic bias and variance like Lemma 1 in closed forms. Thus, in this case, we need to resort to the parametric bootstrap method given in Section 2 to estimate the cMSE.

Appendix

We here give a proof of Lemma 1. Using the results in Ghosh and Maiti (2004), we immediately have $\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta} = \boldsymbol{U}(\boldsymbol{\eta})^{-1}\boldsymbol{s}_m + o_p(m^{-1/2})$. Using this expression, we have

$$E\left[(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t|y_i\right] = \boldsymbol{U}(\boldsymbol{\eta})^{-1}E\left[\boldsymbol{s}_m\boldsymbol{s}_m^t|y_i\right]\boldsymbol{U}(\boldsymbol{\eta})^{-1} + o_p(m^{-1}),$$

where

$$\begin{aligned}
E\left[\boldsymbol{s}_m\boldsymbol{s}_m^t|y_i\right] &= \sum_{j=1}^m E\left[\boldsymbol{D}_j^t\boldsymbol{\Sigma}_j^{-1}\boldsymbol{g}_j\boldsymbol{g}_j^t\boldsymbol{\Sigma}_j^{-1}\boldsymbol{D}_j|y_i\right]\\
&= \sum_{j\neq i}^m E\left[\boldsymbol{D}_j^t\boldsymbol{\Sigma}_j^{-1}\boldsymbol{g}_j\boldsymbol{g}_j^t\boldsymbol{\Sigma}_j^{-1}\boldsymbol{D}_j\right] + \boldsymbol{D}_i^t\boldsymbol{\Sigma}_i^{-1}\boldsymbol{g}_i\boldsymbol{g}_i^t\boldsymbol{\Sigma}_i^{-1}\boldsymbol{D}_i\\
&= \boldsymbol{U}(\boldsymbol{\eta}) + \boldsymbol{D}_i^t\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{g}_i\boldsymbol{g}_i^t - \boldsymbol{\Sigma}_i)\boldsymbol{\Sigma}_i^{-1}\boldsymbol{D}_i,
\end{aligned}$$

since g_j depends only on y_j of $\boldsymbol{Y}$ and $y_1, \ldots, y_m$ are mutually independent. Since $\boldsymbol{U}(\boldsymbol{\eta}) = O(m)$ and $\boldsymbol{D}_i^t\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{g}_i\boldsymbol{g}_i^t - \boldsymbol{\Sigma}_i)\boldsymbol{\Sigma}_i^{-1}\boldsymbol{D}_i = O_p(1)$, we have $E\left[\boldsymbol{s}_m\boldsymbol{s}_m^t|y_i\right] = \boldsymbol{U}(\boldsymbol{\eta}) + O_p(1)$, so that

$$E\left[(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t|y_i\right] = \boldsymbol{U}(\boldsymbol{\eta})^{-1} + o_p(m^{-1}).$$

Next, we evaluate the asymptotical conditional bias of $\widehat{\boldsymbol{\eta}}$, i.e. $E[\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}|y_i]$. By expanding the equation (17) up to second order, we have

$$\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta} = \Big(-\frac{\partial \boldsymbol{s}_m}{\partial \boldsymbol{\eta}}\Big)^{-1}\Big(\boldsymbol{s}_m + \frac{1}{2}\boldsymbol{t} + o_p(1)\Big),$$

where

$$\frac{\partial \boldsymbol{s}_m}{\partial \boldsymbol{\eta}} = \sum_{j=1}^m \Big(\frac{\partial}{\partial \boldsymbol{\eta}^t}\boldsymbol{D}_j^t\boldsymbol{\Sigma}_j^{-1}\Big)\left(\boldsymbol{I}_p \otimes \boldsymbol{g}_j\right) + \sum_{j=1}^m \boldsymbol{D}_j^t\boldsymbol{\Sigma}_j^{-1}\Big(\frac{\partial \boldsymbol{g}_j}{\partial \boldsymbol{\eta}^t}\Big) = \boldsymbol{A}_1 + \boldsymbol{A}_2, \quad \text{(say)}$$

and

$$\boldsymbol{t} = \mathbf{col}_\ell\Big\{(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t\left(\frac{\partial^2 S_{m.\ell}}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}^t}\right)(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})\Big\},$$

for $\boldsymbol{s}_m = (S_{m.1}, \ldots, S_{m.p+1})$. The notation $\mathbf{col}_\ell\{A_\ell\}$ for matrix A_ℓ, $\ell = 1, \ldots, n$ is defined by

$$\mathbf{col}_\ell\{A_\ell\} = (A_1' \ A_2' \ \ldots \ A_n')'$$

Note that $E(\boldsymbol{A}_1) = \boldsymbol{0}$, $\boldsymbol{A}_1 = O_p(m^{1/2})$, $E(\boldsymbol{A}_2) = \boldsymbol{U}(\boldsymbol{\eta})$ and $\boldsymbol{A}_2 - \boldsymbol{U}(\boldsymbol{\eta}) = O_p(m^{1/2})$. Then, we have

$$\begin{aligned}
E[\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}|y_i] &= E\left[\boldsymbol{U}(\boldsymbol{\eta})^{-1}\boldsymbol{A}_1\boldsymbol{U}(\boldsymbol{\eta})^{-1}\boldsymbol{s}_m|y_i\right] + E\left[\boldsymbol{U}(\boldsymbol{\eta})^{-1}\boldsymbol{A}_2\boldsymbol{U}(\boldsymbol{\eta})^{-1}\boldsymbol{s}_m|y_i\right]\\
&\quad + \frac{1}{2}E\left[\boldsymbol{U}(\boldsymbol{\eta})^{-1}\boldsymbol{t}|y_i\right] + o_p(m^{-1})\\
&= \boldsymbol{I}_1(y_i, \boldsymbol{\eta}) + \boldsymbol{I}_2(y_i, \boldsymbol{\eta}) + \frac{1}{2}\boldsymbol{I}_3(y_i, \boldsymbol{\eta}) + o_p(m^{-1}). \quad \text{(say)}
\end{aligned}$$

Since $E(\boldsymbol{g}_j)=0$, $\boldsymbol{g}_1,\ldots,\boldsymbol{g}_m$ are mutually independent, and $\boldsymbol{g}_i$ depends only on y_i of $\boldsymbol{Y}$, we have

$$\boldsymbol{I}_1(y_i,\boldsymbol{\eta})=\sum_{j=1}^m \boldsymbol{U}(\boldsymbol{\eta})^{-1}\Big(\frac{\partial}{\partial\boldsymbol{\eta}^t}\boldsymbol{D}_j^t\boldsymbol{\Sigma}_j^{-1}\Big)E\left[\left(\boldsymbol{I}_p\otimes\boldsymbol{g}_j\right)\boldsymbol{C}_j\boldsymbol{g}_j|y_i\right],$$

where $\boldsymbol{C}_j=\boldsymbol{U}(\boldsymbol{\eta})^{-1}\boldsymbol{D}_j^t\boldsymbol{\Sigma}_j^{-1}(=O(m^{-1}))$. Using the fact that

$$\left(\boldsymbol{I}_p\otimes\boldsymbol{g}_j\right)\boldsymbol{C}_j\boldsymbol{g}_j=\mathbf{vec}\,(\boldsymbol{g}_j\boldsymbol{g}_j^t\boldsymbol{C}_j^t),\quad\text{and}\quad E\left[\mathbf{vec}\,(\boldsymbol{g}_j\boldsymbol{g}_j^t\boldsymbol{C}_j^t)\right]=\mathbf{vec}\,(\boldsymbol{D}_j\boldsymbol{U}(\boldsymbol{\eta})^{-1}),$$

we have

$$\begin{aligned}\boldsymbol{I}_1(y_i,\boldsymbol{\eta})&=\sum_{j=1}^m \boldsymbol{U}(\boldsymbol{\eta})^{-1}\Big(\frac{\partial}{\partial\boldsymbol{\eta}^t}\boldsymbol{D}_j^t\boldsymbol{\Sigma}_j^{-1}\Big)\mathbf{vec}\,(\boldsymbol{D}_j\boldsymbol{U}(\boldsymbol{\eta})^{-1})\\&\quad+\boldsymbol{U}(\boldsymbol{\eta})^{-1}\Big(\frac{\partial}{\partial\boldsymbol{\eta}^t}\boldsymbol{D}_i^t\boldsymbol{\Sigma}_i^{-1}\Big)\mathbf{vec}\,(\boldsymbol{g}_i\boldsymbol{g}_i^t\boldsymbol{C}_i^t-\boldsymbol{D}_j\boldsymbol{U}(\boldsymbol{\eta})^{-1})\\&=\sum_{j=1}^m \boldsymbol{U}(\boldsymbol{\eta})^{-1}\Big(\frac{\partial}{\partial\boldsymbol{\eta}^t}\boldsymbol{D}_j^t\boldsymbol{\Sigma}_j^{-1}\Big)\mathbf{vec}\,(\boldsymbol{D}_j\boldsymbol{U}(\boldsymbol{\eta})^{-1})+o_p(m^{-1}).\end{aligned}\tag{23}$$

Similarly, we have

$$\boldsymbol{I}_2(y_i,\boldsymbol{\eta})=\sum_{j=1}^m \boldsymbol{U}(\boldsymbol{\eta})^{-1}\boldsymbol{D}_j^t\boldsymbol{\Sigma}_j^{-1}E\left[\boldsymbol{h}_j\boldsymbol{C}_j\boldsymbol{g}_j|y_i\right],$$

where

$$\boldsymbol{h}_j\equiv\frac{\partial\boldsymbol{g}_j}{\partial\boldsymbol{\eta}^t}=Q(m_j)\begin{pmatrix}\boldsymbol{x}_j^t & 0\\ \{2(y_j-m_j)+\phi_jQ'(m_j)\}\,\boldsymbol{x}_j^t & -(1+v_2/n_j)(\nu-v_2)^{-2}\end{pmatrix}.$$

Since

$$E\left[\boldsymbol{h}_j\boldsymbol{C}_j\boldsymbol{g}_j\right]=2Q(m_j)\begin{pmatrix}\mathbf{0}^t & 0\\ \boldsymbol{x}_j^t & 0\end{pmatrix}\boldsymbol{C}_j\begin{pmatrix}\mu_{2j}\\ \mu_{3j}\end{pmatrix}=\boldsymbol{E}_j,\qquad\text{(say)}$$

and $\boldsymbol{E}_j=O_p(m^{-1})$, we have

$$\begin{aligned}\boldsymbol{I}_2(y_i,\boldsymbol{\eta})&=\sum_{j=1}^m \boldsymbol{U}(\boldsymbol{\eta})^{-1}\boldsymbol{D}_j^t\boldsymbol{\Sigma}_j^{-1}\boldsymbol{E}_j+\boldsymbol{U}(\boldsymbol{\eta})^{-1}\boldsymbol{D}_i^t\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{h}_i\boldsymbol{C}_i\boldsymbol{g}_i-\boldsymbol{E}_i\right)\\&=\sum_{j=1}^m \boldsymbol{U}(\boldsymbol{\eta})^{-1}\boldsymbol{D}_j^t\boldsymbol{\Sigma}_j^{-1}\boldsymbol{E}_j+o_p(m^{-1}).\end{aligned}\tag{24}$$

For the evaluation of $\boldsymbol{I}_3(y_i)$, we observe that

$$\begin{aligned}E\left[\boldsymbol{t}|y_i\right]&=\mathbf{col}_\ell\Big\{\mathrm{tr}\,E\Big[\Big(\frac{\partial S_{m.\ell}}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^t}\Big)\boldsymbol{U}(\boldsymbol{\eta})^{-1}\boldsymbol{s}_m\boldsymbol{s}_m^t\boldsymbol{U}(\boldsymbol{\eta})^{-1}\Big|y_i\Big]\Big\}+o_p(m^{-1})\\&=\mathbf{col}_\ell\Big\{\mathrm{tr}\sum_{j=1}^m\boldsymbol{K}_{j\ell}\boldsymbol{U}(\boldsymbol{\eta})^{-1}\Big\}+o_p(m^{-1}),\end{aligned}$$

where

$$\boldsymbol{K}_{jl}=2\frac{\partial \boldsymbol{d}_{j\ell}}{\partial \boldsymbol{\eta}}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{D}_j+2(\boldsymbol{I}_p\otimes \boldsymbol{d}_{j\ell})\frac{\partial \boldsymbol{\Sigma}_j^{-1}}{\partial \boldsymbol{\eta}}\boldsymbol{D}_j+(\boldsymbol{I}_p\otimes(\boldsymbol{d}_{j\ell}\boldsymbol{\Sigma}_j^{-1}))\boldsymbol{J}_j$$

for $\boldsymbol{J}_j=[\boldsymbol{G}_{1j}^t,\boldsymbol{G}_{2j}^t]^t$, which are expressed as

$$\boldsymbol{G}_{1j}=-Q(m_i)\boldsymbol{x}_i\otimes\begin{bmatrix} Q'(m_i)\boldsymbol{x}_i^t & 0\\ \{-2Q(m_i)+(Q'(m_i))^2+2\phi_i v_2 Q(m_i)\}\boldsymbol{x}_i^t & -Q'(m_i)(1+v_2/n_i)(\lambda-v_2)^{-2}\end{bmatrix},$$

and

$$\boldsymbol{G}_{2j}=-Q(m_i)\begin{bmatrix} \boldsymbol{0}^t & 0\\ -(1+v_2/n_i)(\lambda-v_2)^{-2}Q'(m_i)\boldsymbol{x}_i^t & 2(1+v_2/n_i)(\lambda-v_2)^{-3}\end{bmatrix}.$$

Then, we have

$$\boldsymbol{I}_3(y_i)=\boldsymbol{U}(\boldsymbol{\eta})^{-1}\mathbf{col}_\ell\Big\{\mathrm{tr}\sum_{j=1}^m\boldsymbol{K}_{j\ell}\boldsymbol{U}(\boldsymbol{\eta})^{-1}\Big\}+o_p(m^{-1}). \tag{25}$$

Combining (23), (24) and (25), one gets the expressions of $\boldsymbol{a}_1(y_i,\boldsymbol{\eta})$ and $\boldsymbol{a}_2(y_i,\boldsymbol{\eta})$ given by

$$\begin{aligned}\boldsymbol{a}_1(\boldsymbol{\eta})=&\sum_{j=1}^m\Big(\frac{\partial}{\partial\boldsymbol{\eta}^t}\boldsymbol{D}_j^t\boldsymbol{\Sigma}_j^{-1}\Big)\mathbf{vec}\,(\boldsymbol{D}_j\boldsymbol{U}(\boldsymbol{\eta})^{-1})+\sum_{j=1}^m\boldsymbol{D}_j^t\boldsymbol{\Sigma}_j^{-1}\boldsymbol{E}_j,\\ \boldsymbol{a}_2(\boldsymbol{\eta})=&\mathbf{col}_\ell\Big\{\mathrm{tr}\,\Big[\sum_{j=1}^m\boldsymbol{K}_{j\ell}\boldsymbol{U}(\boldsymbol{\eta})^{-1}\Big]\Big\},\end{aligned}$$

which completes the proof.

Acknowledgments

The second author acknowledges support from Grant-in-Aid for Scientific Research (23243039 and 26330036), Japan.

References

[1] Booth, J. S. and Hobert, P. (1998). Standard errors of prediction in generalized linear mixed models. *Journal of the American Statistical Association.* **93**, 262 - 272.

[2] 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.

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

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

[5] Godambe, V. P. and Thompson, M. E. (1989). An extension of quasi-likelihood estimation (with Discussion). *Journal of Statistical Planning and Inference.* **22**, 137 - 152.

[6] Ghosh, M. and Maiti, T. (2004). Small-area estimation based on natural exponential family quadratic variance function models and survey weights. *Biometrika.* **91**, 95 - 112.

[7] Ghosh, M. and Maiti, T. (2008). Empirical Bayes Confidence Intervals for Means of Natural Exponential Family-Quadratic Variance Function Distributions with Application to Small Area Estimation. *Scandinavian Journal of Statistics.* **35**, 484 - 495.

[8] Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science*, **9**, 55 - 93.

[9] Hall, P. and Maiti, T. (2006). On parametric bootstrap methods for small area prediction. *Journal of Royal Statistical Society: Series B.* **68**, 221 - 238.

[10] Kubokawa, T., Hasukawa, M. and Takahashi, K. (2014). On measuring uncertainty of benchmarked predictors with application to disease risk estimate. *Scandinavian Journal of Statistics*, **41**, 394 - 413.

[11] Morris, C. (1982). Natural exponential families with quadratic variance functions. *The Annals of Statistics.* **10**, 65 - 80.

[12] Morris, C. (1983). Natural exponential families with quadratic variance functions: Statistical theory. *The Annals of Statistics*, **11**, 515 - 529.

[13] Prasad, N. and Rao, J. N. K. (1990). The estimation of mean-squared errors of small-area estimators. *Journal of the American Statistical Association*, **90**, 758 - 766.

[14] Rao, J.N.K. (2003). *Small Area Estimation.* Wiley.

[15] Torabi, M. and Rao, J.N.K. (2013). Estimation of mean squared error of model-based estimators of small area means under a nested error linear regression model. *Journal of Multivariate Analysis.* **117**, 76 - 87.