CIRJE-F-861

On Measuring Uncertainty of Benchmarked Predictors with Application to Disease Risk Estimate

Tatsuya Kubokawa
University of Tokyo

Mana Hasukawa
Graduate School of Economics, University of Tokyo

Kunihiko Takahashi
National Institute of Public Health

September 2012

On Measuring Uncertainty of Benchmarked Predictors with Application to Disease Risk Estimate

Tatsuya Kubokawa*, Mana Hasukawa†and Kunihiko Takahashi‡
University of Tokyo and NIPH

September 12, 2012

Abstract

Empirical Bayes (EB) estimates in general linear mixed models are useful for the small area estimation in the sense of increasing precision of estimation of small area means. However, one potential difficulty of EB is that the overall estimate for a larger geographical area based on a (weighted) sum of EB estimates is not necessarily identical to the corresponding direct estimate like the overall sample mean. Another difficulty is that EB estimates yield over-shrinking, which results in the sampling variance smaller than the posterior variance. One way to fix these problems is the benchmarking approach based on the constrained empirical Bayes (CEB) estimators, which satisfy the constraints that the aggregated mean and variance are identical to the requested values of mean and variance. In this paper, we treat the general mixed models, derive asymptotic approximations of the mean squared error (MSE) of CEB and provide second-order unbiased estimators of MSE based on the parametric bootstrap method. These results are applied to natural exponential families with quadratic variance functions (NEF-QVF). As a specific example, the Poisson-gamma model is dealt with, and it is illustrated that the CEB estimates and their MSE estimates work well through real mortality data.

Key words and phrases: Asymptotic approximation, benchmarking, best linear unbiased predictor, Binomial-beta model, constrained Bayes, empirical Bayes, estimating equation, generalized linear mixed model, mean squared error, natural exponential family, parametric bootstrap, mortality rates, Poisson-gamma model, second-order approximation, small area estimation.

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

†Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, E-Mail: hs.mana21@gmail.com

‡National Institute of Public Health, 2-3-6 Minami, Wako, Saitama 351-0197, Japan. E-Mail: kunihiko@niph.go.jp

1 Introduction

The usefulness of the empirical Bayes (EB) estimators in various mixed models has been recognized in the literature. Of these, EB in Poisson-gamma models has been used for disease mapping and estimation of mortality rates by Tsutakawa (1985, 88), Tsutakawa, Shoop and Marienfeld (1985), Clayton and Kaldor (1987), Tango (1988) and Manton *et al*. (1989), and their extensions to generalized linear mixed models have been studied by Breslow and Clayton (1993), Breslow and Lin (1995), Sarkar and Ghosh (1998) and others. To measure uncertainty of EB, Ghosh and Maiti (2004) and Crescenzi, Ghosh and Maiti (2005) obtained asymptotic approximations of mean squared error (MSE) of EB and their asymptotically unbiased estimators in natural exponential families with quadratic variance functions (NEF-QVF). The confidence intervals based on EB were constructed by Ghosh and Maiti (2008) in NEF-QVF.

A good application of EB is the small area estimation, where direct estimates like sample means for small areas have unacceptable estimation errors because sample sizes of small areas are small. The empirical Bayes procedures, which is called the empirical best linear unbiased predictor (EBLUP) in the context of linear mixed models, are alternative methods to provide stable estimates with higher precisions by borrowing data in the surrounding areas. However, one potential difficulty of EB is that the overall estimate for a larger geographical area, which is constructed by a (weighted) sum of EB estimates of individual small areas, is not necessarily equal to the corresponding direct estimate like the overall sample mean. Another difficulty is that EB estimates yield over-shrinking as illustrated in figures in Section 5.2. In fact, Louis (1984) and Ghosh (1992) pointed out and showed that the sampling variance of EB is smaller than the posterior variance. One way to solve these problems is the benchmarking approach, which modifies EB so that one gets the same (weighted) aggregate mean and/or variance for the larger geographical area. Ghosh (1992) suggested the constrained Bayes (CB) estimator and the constrained empirical Bayes (CEB) estimator which satisfy the constraints that the aggregated mean and variance are identical to the mean and variance of the posterior distribution, and Datta, Ghosh, Steorts and Maples (2011) and Kubokawa (2012) gave some extensions. Frey and Cressie (2003) derived CEB in Poisson-gamma models.

Since the sample variance of EB is smaller than the posterior variance, CEB modifies EB so that its sample variance is identical to the posterior variance. However, the usefulness and purpose of EB is that EB gives stable estimates with higher precision of estimation. Then we have a concern whether CEB may be against this purpose. Thus, it is quite interesting and important to assess the mean squared error (MSE) of CEB. In this paper, we address this issue for the general mixed models with applications to estimation of mortality rates.

In Section 2, we first give unified results about MSE of EB and estimation of MSE in general mixed models. Extensions to the benchmarking problems in general mixed models are developed in Section 3. Asymptotic approximations of MSE for CEB are derived and second-order unbiased estimators of MSE for CEB are provided by the parametric bootstrap method from Butar and Lahiri (2003). When the variance constraint is the posterior variance, it is shown that MSE of CEB is larger than MSE of EB in the first order approximation. To modify this property, we suggest some modification of the variance constraint. In Section 4, we apply the general results in Section 3 to natural exponential families with quadratic variance functions (NEF-QVF). NEF-QVF distributions were discussed by Morris (1982, 83), and MSE and estimation of MSE for EB were studied by Ghosh and Maiti (2004). We make use of their tools and results to construct estimation of MSE of CEB. Applications to Poisson-gamma and binomial-beta mixture models are given in Section 5, and benchmarking SMR in estimation of mortality rates is investigated in Section 5.2 through real data of the mortality from stomach cancer of females in 92 cities or towns in Saitama prefecture, which is next to Tokyo, for five years from 1995 to 1999. Finally, the concluding remarks are given in Section 6.

2 Unified Results in General Mixed Model

2.1 MSE of empirical Byes estimator

Let $\boldsymbol{y}$ be an N dimensional vector of observable random variables, and let $\boldsymbol{\theta}$ be a p dimensional 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 $\boldsymbol{y}$ and $\boldsymbol{\theta}$. The conditional probability density (or mass) function of $\boldsymbol{y}$ given $(\boldsymbol{\theta},\boldsymbol{\eta})$ is denoted by $f(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\eta})$, and the conditional probability density (or mass) function of $\boldsymbol{\theta}$ given $\boldsymbol{\eta}$ is denoted by $\pi(\boldsymbol{\theta}|\boldsymbol{\eta})$, namely,

$$\begin{aligned} \boldsymbol{y}|(\boldsymbol{\theta},\boldsymbol{\eta}) \sim& f(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\eta}),\\ \boldsymbol{\theta}|\boldsymbol{\eta} \sim& \pi(\boldsymbol{\theta}|\boldsymbol{\eta}). \end{aligned} \tag{2.1}$$

This expresses 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 $\boldsymbol{y}$ for given $\boldsymbol{\eta}$ and the conditional (or posterior) density function of $\boldsymbol{\theta}$ given $(\boldsymbol{y},\boldsymbol{\eta})$ are given by

$$\begin{aligned} m_\pi(\boldsymbol{y}|\boldsymbol{\eta}) =& \int f(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\eta})\pi(\boldsymbol{\theta}|\boldsymbol{\eta})\mathrm{d}\boldsymbol{\theta},\\ \pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{\eta}) =& f(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\eta})\pi(\boldsymbol{\theta}|\boldsymbol{\eta})/m_\pi(\boldsymbol{y}|\boldsymbol{\eta}), \end{aligned} \tag{2.2}$$

and we use the same notations in the discrete case. Then, we consider the problem of predicting the scalar quantity $\xi(\boldsymbol{\theta},\boldsymbol{\eta})$ where a predictor $\widehat{\xi}(\boldsymbol{y})$ is evaluated in terms of the mean squared error $MSE(\boldsymbol{\eta},\widehat{\xi}) = E[(\widehat{\xi}(\boldsymbol{y}) - \xi(\boldsymbol{\theta},\boldsymbol{\eta}))^2]$. When $\boldsymbol{\eta}$ is known, the best predictor of $\xi(\boldsymbol{\theta},\boldsymbol{\eta})$ in the sense of minimizing the MSE is the conditional expectation given by

$$\widehat{\xi}(\boldsymbol{y},\boldsymbol{\eta}) = E[\xi(\boldsymbol{\theta},\boldsymbol{\eta})|\boldsymbol{y}],$$

which is the Bayes estimator in the Bayesian context. Since $\boldsymbol{\eta}$ is unknown in this paper, we need to estimate $\boldsymbol{\eta}$ from the marginal density function $m_\pi(\boldsymbol{y}|\boldsymbol{\eta})$. Substituting an estimator $\widehat{\boldsymbol{\eta}}$ into $\widehat{\xi}(\boldsymbol{y},\boldsymbol{\eta})$, we get the empirical Bayes estimator (EB) $\widehat{\xi}(\boldsymbol{y},\widehat{\boldsymbol{\eta}})$.

We next derive a second-order approximation of MSE of $\widehat{\xi}(\boldsymbol{y},\widehat{\boldsymbol{\eta}})$ and to provide a second-order unbiased estimator of MSE based on the parametric bootstrap method. We assume that there exists a consistent estimator $\widehat{\boldsymbol{\eta}}$ of $\boldsymbol{\eta}$ satisfying the following condition:

(A1) The dimension q of $\boldsymbol{\eta}$ is bounded. Estimator $\widehat{\boldsymbol{\eta}}$ satisfies that $\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta}=O_p(N^{-1/2})$ and $E[\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta}]=O(N^{-1})$.

(A2) The Bayes estimator $\widehat{\xi}(\boldsymbol{y},\boldsymbol{\eta})$ is continuously differentiable with respect to $\boldsymbol{\eta}$, and $\partial\widehat{\xi}(\boldsymbol{y},\boldsymbol{\eta})/\partial\eta_i=O_p(1)$ for large N and $i=1,\ldots,q$.

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

$$g_1(\boldsymbol{\eta}) =E[Var(\xi(\boldsymbol{\theta},\boldsymbol{\eta})|\boldsymbol{y})], \tag{2.3}$$

$$g_2(\boldsymbol{\eta}) =E\Big[\{(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})^t\frac{\partial\widehat{\xi}(\boldsymbol{y},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\}^2\Big]. \tag{2.4}$$

Theorem 2.1 *Assume conditions* (A1) *and* (A2) *with* $g_1(\boldsymbol{\eta})=O(1)$*. Then, the MSE of* $\widehat{\xi}(\boldsymbol{y},\widehat{\boldsymbol{\eta}})$ *is approximated as*

$$MSE(\boldsymbol{\eta},\widehat{\xi}(\boldsymbol{y},\widehat{\boldsymbol{\eta}}))=g_1(\boldsymbol{\eta})+g_2(\boldsymbol{\eta})+o(N^{-1}), \tag{2.5}$$

where $g_2(\boldsymbol{\eta})=O(N^{-1})$.

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

$$\begin{aligned}MSE(\boldsymbol{\eta},\widehat{\xi}(\boldsymbol{y},\widehat{\boldsymbol{\eta}})) =&E[\{\xi(\boldsymbol{\theta},\boldsymbol{\eta})-\widehat{\xi}(\boldsymbol{y},\boldsymbol{\eta})+\widehat{\xi}(\boldsymbol{y},\boldsymbol{\eta})-\widehat{\xi}(\boldsymbol{y},\widehat{\boldsymbol{\eta}})\}^2]\\ =&E[\{\xi(\boldsymbol{\theta},\boldsymbol{\eta})-\widehat{\xi}(\boldsymbol{y},\boldsymbol{\eta})\}^2]+E[\{\widehat{\xi}(\boldsymbol{y},\widehat{\boldsymbol{\eta}})-\widehat{\xi}(\boldsymbol{y},\boldsymbol{\eta})\}^2],\end{aligned} \tag{2.6}$$

and that $E[\{\xi(\boldsymbol{\theta},\boldsymbol{\eta})-\widehat{\xi}(\boldsymbol{y},\boldsymbol{\eta})\}^2]=E[Var(\xi(\boldsymbol{\theta},\boldsymbol{\eta})|\boldsymbol{y})]=g_1(\boldsymbol{\eta})$. It is noted that

$$\widehat{\xi}(\boldsymbol{y},\widehat{\boldsymbol{\eta}})=\widehat{\xi}(\boldsymbol{y},\boldsymbol{\eta})+(\frac{\partial\widehat{\xi}(\boldsymbol{y},\boldsymbol{\eta}^*)}{\partial\boldsymbol{\eta}})^t(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta}), \tag{2.7}$$

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

$$E[\{\widehat{\xi}(\boldsymbol{y},\widehat{\boldsymbol{\eta}})-\widehat{\xi}(\boldsymbol{y},\boldsymbol{\eta})\}^2]=E\Big[\{(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})^t\frac{\partial\widehat{\xi}(\boldsymbol{y},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\}^2\Big]+o(N^{-1}),$$

which shows Theorem 2.1. ■

Remark 2.1 The moment conditions stated in (A1) are conditions that we have to impose at least for $\widehat{\boldsymbol{\eta}}$. If $\widehat{\boldsymbol{\eta}}$ were expanded as

$$\widehat{\boldsymbol{\eta}}=\boldsymbol{\eta}+\boldsymbol{\eta}^\dagger+\boldsymbol{\eta}^{\dagger\dagger}+O_p(N^{-3/2}),$$

where $\boldsymbol{\eta}^\dagger=O_p(N^{-1/2})$, $E[\boldsymbol{\eta}^\dagger]=\mathbf{0}$ and $\boldsymbol{\eta}^{\dagger\dagger}=O_p(N^{-1})$, one could establish higher order accuracy for the approximations of MSE. As shown in Kubokawa (2011), such expansions are available for ML, REML and some specific estimators in normal linear mixed models. It may be, however, harder to get such an expansion in non-normal mixed models.

2.2 Estimation of MSE of EB

We now derive two kinds of second-order unbiased estimators of MSE for EB from Theorem 2.1; One is based on the Taylor series expansion and the other is based on the parametric bootstrap method. Assume the following condition:

(A3) $g_1(\boldsymbol{\eta})$ is twice continuously differentiable, and $g_1(\boldsymbol{\eta}) = O(1)$, $\partial g_1(\boldsymbol{\eta})/\partial\eta_i = O_p(1)$ and $\partial^2 g_1(\boldsymbol{\eta})/\partial\eta_i\partial\eta_j = O_p(1)$ for large N and $i, j = 1, \ldots, q$.

Under condition (A3), it is noted that

$$g_1(\widehat{\boldsymbol{\eta}}) = g_1(\boldsymbol{\eta}) + \big(\frac{\partial g_1(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\big)^t(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) + \frac{1}{2}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t \frac{\partial^2 g_1(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^t}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}).$$

This implies that $E[g_1(\widehat{\boldsymbol{\eta}})] = g_1(\boldsymbol{\eta}) + g_{11}(\boldsymbol{\eta}) + g_{12}(\boldsymbol{\eta}) + o(N^{-1})$, where

$$g_{11}(\boldsymbol{\eta}) = E\Big[\big(\frac{\partial g_1(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\big)^t(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})\Big], \tag{2.8}$$

$$g_{12}(\boldsymbol{\eta}) = \frac{1}{2}E\Big[(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t \frac{\partial^2 g_1(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^t}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})\Big]. \tag{2.9}$$

Noting that $g_{11}(\boldsymbol{\eta}) = O(N^{-1})$, $g_{12}(\boldsymbol{\eta}) = O(N^{-1})$ and $g_2(\boldsymbol{\eta}) = O(N^{-1})$, and that these functions are continuous with respect to $\boldsymbol{\eta}$, we can see that $E[g_1(\widehat{\boldsymbol{\eta}}) - g_{11}(\widehat{\boldsymbol{\eta}}) - g_{12}(\widehat{\boldsymbol{\eta}})] = g_1(\boldsymbol{\eta}) + o(N^{-1})$ and $E[g_2(\widehat{\boldsymbol{\eta}})] = g_2(\boldsymbol{\eta}) + o(N^{-1})$.

Theorem 2.2 *Assume conditions* (A1), (A2) *and* (A3). *Then, a second-order unbiased estimator of* $MSE(\boldsymbol{\eta}, \widehat{\xi}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}))$ *is given by*

$$mse(\boldsymbol{y}, \widehat{\xi}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}})) = g_1(\widehat{\boldsymbol{\eta}}) - g_{11}(\widehat{\boldsymbol{\eta}}) - g_{12}(\widehat{\boldsymbol{\eta}}) + g_2(\widehat{\boldsymbol{\eta}}), \tag{2.10}$$

namely, $E[mse(\boldsymbol{y}, \widehat{\xi}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}))] = MSE(\boldsymbol{\eta}, \widehat{\xi}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}})) + o(N^{-1})$.

For more complicated models with large q, it may be harder to compute the moments $E[(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})^t(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})]$, $E[\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta}]$ and differentiations of $\widehat{\xi}(\boldsymbol{y}, \boldsymbol{\eta})$ and $g_1(\boldsymbol{\eta})$. Instead of Theorem 2.2, we use the parametric bootstrap method using the same arguments as in Butar and Lahiri (2003). Consider the following model based on the parametric bootstrap:

$$\begin{aligned} \boldsymbol{y}^*|(\boldsymbol{\theta}^*, \widehat{\boldsymbol{\eta}}) \sim& f(\boldsymbol{y}^*|\boldsymbol{\theta}^*, \widehat{\boldsymbol{\eta}}), \\ \boldsymbol{\theta}^*|\widehat{\boldsymbol{\eta}} \sim& \pi(\boldsymbol{\theta}^*|\widehat{\boldsymbol{\eta}}). \end{aligned} \tag{2.11}$$

Let $\widehat{\boldsymbol{\eta}}^*$ be an estimator based on $\boldsymbol{y}^*$, where the calculation of $\widehat{\boldsymbol{\eta}}^*$ is the same as that of $\widehat{\boldsymbol{\eta}}$ except that $\widehat{\boldsymbol{\eta}}^*$ is calculated based on $\boldsymbol{y}^*$ instead of $\boldsymbol{y}$. Since $E[g_1(\widehat{\boldsymbol{\eta}})] = g_1(\boldsymbol{\eta}) + g_{11}(\boldsymbol{\eta}) + g_{12}(\boldsymbol{\eta}) + o(N^{-1})$, it is seen that

$$E^*[g_1(\widehat{\boldsymbol{\eta}}^*)] = g_1(\widehat{\boldsymbol{\eta}}) + g_{11}(\widehat{\boldsymbol{\eta}}) + g_{12}(\widehat{\boldsymbol{\eta}}) + o_p(N^{-1}),$$

where $E^*[\cdot]$ denotes expectation with respect to $\boldsymbol{y}^*$. Thus, we have

$$E[E^*[g_1(\widehat{\boldsymbol{\eta}}^*)]] = E[2g_1(\widehat{\boldsymbol{\eta}})] - g_1(\boldsymbol{\eta}) + o(N^{-1}),$$

or

$$g_1(\boldsymbol{\eta}) = E[2g_1(\widehat{\boldsymbol{\eta}}) - E^*[g_1(\widehat{\boldsymbol{\eta}}^*)]] + o(N^{-1}).$$

To estimate the second term in the r.h.s. of (2.6), let

$$g_2^*(\widehat{\boldsymbol{\eta}}) = E^*[\{\widehat{\xi}(\boldsymbol{y}^*, \widehat{\boldsymbol{\eta}}^*) - \widehat{\xi}(\boldsymbol{y}^*, \widehat{\boldsymbol{\eta}})\}^2].$$

Since $g_2(\boldsymbol{\eta}) = O(N^{-1})$, it is seen that $E[g_2^*(\widehat{\boldsymbol{\eta}})] = E[g_2(\widehat{\boldsymbol{\eta}}) + o_p(N^{-1})] = g_2(\boldsymbol{\eta}) + o(N^{-1})$, and we get the following estimator.

Theorem 2.3 *Assume conditions* (A1), (A2) *and* (A3). *Then, a second-order unbiased estimator of* $MSE(\boldsymbol{\eta}, \widehat{\xi}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}))$ *is given by*

$$mse^*(\boldsymbol{y}, \widehat{\xi}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}})) = 2g_1(\widehat{\boldsymbol{\eta}}) - E^*[g_1(\widehat{\boldsymbol{\eta}}^*)] + g_2^*(\widehat{\boldsymbol{\eta}}), \tag{2.12}$$

namely, $E[mse^*(\boldsymbol{y}, \widehat{\xi}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}))] = MSE(\boldsymbol{\eta}, \widehat{\xi}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}})) + o(N^{-1})$.

3 Extensions to Benchmarking Problems

3.1 Benchmarking problems

Consider the case that the total population is divided into K small areas or small domains in the model (2.1). For $i = 1, \ldots, K$, let $\xi_i = \xi_i(\boldsymbol{\theta}, \boldsymbol{\eta})$ be a quantity we want to predict. The best predictor of $\xi_i(\boldsymbol{\theta}, \boldsymbol{\eta})$ relative to the squared error loss is given by $\widehat{\xi}_i(\boldsymbol{y}, \boldsymbol{\eta}) = E[\xi_i(\boldsymbol{\theta}, \boldsymbol{\eta})|\boldsymbol{y}]$. Let w_i's be nonnegative constants such that $\sum_{i=1}^K w_i = 1$. Then, Louis (1984) showed that

$$(1)\ \sum_{i=1}^K w_i \widehat{\xi}_i(\boldsymbol{y}, \boldsymbol{\eta}) = \sum_{i=1}^K w_i E[\xi_i(\boldsymbol{\theta}, \boldsymbol{\eta})|\boldsymbol{y}],$$

$$(2)\ \sum_{i=1}^K w_i \{\widehat{\xi}_i(\boldsymbol{y}, \boldsymbol{\eta}) - \overline{\widehat{\xi}}_w(\boldsymbol{y}, \boldsymbol{\eta})\}^2 \leq \sum_{i=1}^K w_i E[\{\xi_i(\boldsymbol{\theta}, \boldsymbol{\eta}) - \overline{\xi}_w(\boldsymbol{\theta}, \boldsymbol{\eta})\}^2|\boldsymbol{y}],$$

where $\overline{\widehat{\xi}}_w(\boldsymbol{y}, \boldsymbol{\eta}) = \sum_{i=1}^K w_i \widehat{\xi}_i(\boldsymbol{y}, \boldsymbol{\eta})$ and $\overline{\xi}_w(\boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{i=1}^K w_i \xi_i(\boldsymbol{\theta}, \boldsymbol{\eta})$. In fact, the inequality (2) follows from the fact that

$$\begin{aligned} \sum_{i=1}^K w_i E[\{\xi_i(\boldsymbol{\theta}, \boldsymbol{\eta}) - \overline{\xi}_w(\boldsymbol{\theta}, \boldsymbol{\eta})\}^2|\boldsymbol{y}] =& \sum_{i=1}^K w_i \{\widehat{\xi}_i(\boldsymbol{y}, \boldsymbol{\eta}) - \overline{\widehat{\xi}}_w(\boldsymbol{y}, \boldsymbol{\eta})\}^2 \\ &+ \sum_{i=1}^K w_i Var(\xi_i(\boldsymbol{\theta}, \boldsymbol{\eta}) - \overline{\xi}_w(\boldsymbol{\theta}, \boldsymbol{\eta})|\boldsymbol{y}). \end{aligned} \tag{3.1}$$

Louis (1984) pointed out that the inequality was due to over-shrinking, and Louis (1984) and Ghosh (1992) derived constrained Bayes estimators $\widehat{\xi}^{CB}(\boldsymbol{y}, \boldsymbol{\eta})$'s which satisfy the constraint (1) and

$$\sum_{i=1}^K w_i \Big\{\widehat{\xi}_i^{CB}(\boldsymbol{y}, \boldsymbol{\eta}) - \sum_{j=1}^K w_j \widehat{\xi}_j^{CB}(\boldsymbol{y}, \boldsymbol{\eta})\Big\}^2 = \sum_{i=1}^K w_i E[\{\xi_i(\boldsymbol{\theta}, \boldsymbol{\eta}) - \overline{\xi}_w(\boldsymbol{\theta}, \boldsymbol{\eta})\}^2|\boldsymbol{y}].$$

It is interesting to investigate whether such constrained estimators have still smaller MSE. For the purpose, we assume the following condition:

(A4) $K/N \to \gamma$, $0 < \gamma \leq 1$, $\sum_{i=1}^K w_i^2 = O(N^{-1})$, $\sum_{i=1}^K w_i\{\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta}) - E[\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta})]\} = O_p(N^{-1/2})$ and $Var(\xi_i(\boldsymbol{\theta},\boldsymbol{\eta})|\boldsymbol{y}) = O_p(1)$.

In typical applications, we consider the case that $N = K = p$, $w_i = n_i/\sum_{j=1}^K n_j$ for sample sizes n_i, $\xi_i(\boldsymbol{\theta},\boldsymbol{\eta}) = \xi_i(\theta_i,\boldsymbol{\eta})$ for $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_K)^t$, $\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta}) = \widehat{\xi}_i(y_i,\boldsymbol{\eta})$ for $\boldsymbol{y} = (y_1,\ldots,y_K)^t$, and $y_1,\ldots,y_K$ are mutually independently distributed. In this case, condition (A4) is satisfied.

In this section, we consider the following constraints for predictors $\delta_i(\boldsymbol{y})$ of $\xi_i(\boldsymbol{\theta},\boldsymbol{\eta})$ and estimator $\widehat{\boldsymbol{\eta}}$:

(C1) $\sum_{i=1}^K w_i\delta_i(\boldsymbol{y}) = \sum_{i=1}^K w_i\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}}) + \Delta_m(\boldsymbol{y})$, where $\Delta_m(\boldsymbol{y}) = O_p(N^{-1/2})$.

(C2) $\sum_{i=1}^K w_i\{\delta_i(\boldsymbol{y}) - \overline{\delta}_w(\boldsymbol{y})\}^2 = \sum_{i=1}^K w_i\{\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}}) - \overline{\widehat{\xi}}_w(\boldsymbol{y},\widehat{\boldsymbol{\eta}})\}^2 + \Delta_v(\boldsymbol{y})$, where $\overline{\delta}_w(\boldsymbol{y}) = \sum_{i=1}^K w_i\delta_i(\boldsymbol{y})$ and $\Delta_v(\boldsymbol{y}) = O_p(N^{-r})$ for $r \geq 0$.

We call (C1) and (C2) mean and variance constraints, respectively. For (C1), we can treat the case that $\Delta_m(\boldsymbol{y}) = 0$ and $\Delta_m(\boldsymbol{y}) = \sum_{i=1}^K w_i\{\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{0}) - \widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}})\}$ with $E[\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta})] - E[\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{0})] = O(N^{-1/2})$, both of which satisfy that $\Delta_m(\boldsymbol{y}) = O_p(N^{-1/2})$ under (A4). For (C2), an example we treat is

$$\Delta_v(\boldsymbol{y}) = N^{-r}\sum_{i=1}^K w_i Var(\xi_i(\boldsymbol{\theta},\widehat{\boldsymbol{\eta}}) - \overline{\xi}_w(\boldsymbol{\theta},\widehat{\boldsymbol{\eta}})|\boldsymbol{y}). \tag{3.2}$$

In the case of $r = 0$, it follows from (3.1) that $\Delta_v(\boldsymbol{y})$ in (3.2) yields the constraint

$$\sum_{i=1}^K w_i E[\{\delta_i(\boldsymbol{y}) - \overline{\delta}_w(\boldsymbol{y})\}^2|\boldsymbol{y}] = \sum_{i=1}^K w_i E[\{\xi_i(\boldsymbol{\theta},\widehat{\boldsymbol{\eta}}) - \overline{\xi}_w(\boldsymbol{\theta},\widehat{\boldsymbol{\eta}})\}^2|\boldsymbol{y}], \tag{3.3}$$

which corresponds to the constraint treated by Ghosh (1992) and Datta *et al.* (2011).

For benchmarking under constraints (C1) and (C2), Ghosh (1992), Frey and Cressie (2003) and Datta *et al.* (2011) derived the constrained Bayes estimator in the sense of minimizing the conditional MSE $E[\{\delta_i(\boldsymbol{y}) - \xi_i(\boldsymbol{\theta},\boldsymbol{\eta})\}^2|\boldsymbol{y}]$ subject to constraints (C1) and (C2). We call here it the constrained Bayes estimator (CB). Substituting an estimator $\widehat{\boldsymbol{\eta}}$ into CB, we get the constrained empirical Bayes estimator (CEB) given by

$$\delta_i^{CEB}(\boldsymbol{y},\widehat{\boldsymbol{\eta}}) = \widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}}) + \{a_B(\boldsymbol{y}) - 1\}\{\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}}) - \overline{\widehat{\xi}}_w(\boldsymbol{y},\widehat{\boldsymbol{\eta}})\} + \Delta_m(\boldsymbol{y}), \tag{3.4}$$

where

$$\{a_B(\boldsymbol{y})\}^2 = 1 + \frac{\Delta_v(\boldsymbol{y})}{\sum_{i=1}^K w_i\{\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}}) - \overline{\widehat{\xi}}_w(\boldsymbol{y},\widehat{\boldsymbol{\eta}})\}^2}. \tag{3.5}$$

3.2 MSE of constrained empirical Bayes estimator (CEB)

We now evaluate MSE of CEB, which is decomposed as

$$\begin{aligned}MSE(\boldsymbol{\eta},\delta_i^{CEB}) =&E[\{\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}})-\xi_i\}^2]\\&+E[\{(a_B(\boldsymbol{y})-1)(\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}})-\overline{\widehat{\xi}}_w(\boldsymbol{y},\widehat{\boldsymbol{\eta}}))+\Delta_m(\boldsymbol{y})\}^2]\\&+2E[\{\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}})-\xi_i\}\{(a_B(\boldsymbol{y})-1)(\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}})-\overline{\widehat{\xi}}_w(\boldsymbol{y},\widehat{\boldsymbol{\eta}}))+\Delta_m(\boldsymbol{y})\}]\\=&I_1(\boldsymbol{\eta})+I_2(\boldsymbol{\eta})+2I_3(\boldsymbol{\eta}).\quad\text{(say)}\end{aligned}\tag{3.6}$$

Since $I_1(\boldsymbol{\eta})=MSE(\boldsymbol{\eta},\widehat{\xi}_i)$, the second order approximation of $I_1(\boldsymbol{\eta})$ is $I_1(\boldsymbol{\eta})=g_1(\boldsymbol{\eta})+g_2(\boldsymbol{\eta})+o(N^{-1})$ as shown in Theorem 2.1. To evaluate I_2 and I_3, we treat the mean constraint and the mean-variance constraint separately.

[1] Mean Constraint. The mean constraint (C1) corresponds to the case of $a_B(\boldsymbol{y})=1$ in (3.4), namely, CEB is written as

$$\delta_i^{CEB}(\boldsymbol{y},\widehat{\boldsymbol{\eta}})=\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}})+\Delta_m(\boldsymbol{y}),\tag{3.7}$$

which implies that $I_2(\boldsymbol{\eta})=E[\{\Delta_m(\boldsymbol{y})\}^2]$ and $I_3(\boldsymbol{\eta})=E[\{\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}})-\xi_i(\boldsymbol{y},\boldsymbol{\eta})\}\Delta_m(\boldsymbol{y})]$. Since $\Delta_m(\boldsymbol{y})=O_p(N^{-1/2})$, it is seen that $I_2(\boldsymbol{\eta})=O(N^{-1})$, and from (2.7) it is observed that $I_3(\boldsymbol{\eta})=\overline{I}_3(\boldsymbol{\eta})+O(N^{-3/2})$, where

$$\overline{I}_3(\boldsymbol{\eta})=E\Big[\Big(\frac{\partial\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\Big)^t(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})\Delta_m(\boldsymbol{y})\Big]=O(N^{-1}).$$

Proposition 3.1 *Assume conditions* (A1)-(A4)*. Then, CEB subject to the mean constraint* (C1) *is given by* (3.7)*, and the MSE is approximated as*

$$\begin{aligned}MSE(\boldsymbol{\eta},\delta_i^{CEB}(\boldsymbol{y},\widehat{\boldsymbol{\eta}}))=&MSE(\boldsymbol{\eta},\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}}))+E[\{\Delta_m(\boldsymbol{y})\}^2]\\&+2E\Big[\Big(\frac{\partial\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\Big)^t(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})\Delta_m(\boldsymbol{y})\Big]+o(N^{-1}).\end{aligned}$$

This proposition implies that in the mean constraint, the difference in MSE between CEB $\delta_i^{CEB}(\boldsymbol{y},\widehat{\boldsymbol{\eta}})$ and the non-constrained EB $\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}})$ appears in the second-order term.

[2] Mean-Variance Constraint. In this case, the asymptotic property of MSE of CEB given in (3.4) depends on r, the order of $\Delta_v(\boldsymbol{y})$. We first evaluate $a_B(\boldsymbol{y})$. For the purpose, we assume the following condition:

(A5) $\sum_{i=1}^K w_i\{(\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta})-\overline{\widehat{\xi}}_w(\boldsymbol{y},\boldsymbol{\eta}))^2-E[(\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta})-\overline{\widehat{\xi}}_w(\boldsymbol{y},\boldsymbol{\eta}))^2]\}=O_p(N^{-1/2})$.

Lemma 3.1 *Assume conditions* (A1)-(A5)*. Then,*

$$\sum_{i=1}^K w_i\{\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}})-\overline{\widehat{\xi}}_w(\boldsymbol{y},\widehat{\boldsymbol{\eta}})\}^2=\nu_0+\nu_1,\tag{3.8}$$

where

$$\nu_0 = \sum_{i=1}^{K} w_i E\big[\big\{\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta}) - \overline{\widehat{\xi}}_w(\boldsymbol{y},\boldsymbol{\eta})\big\}^2\big],$$

and ν_1 *is a function with* $O_p(N^{-1/2})$*. Also,*

$$a_B(\boldsymbol{y}) - 1 = A_B(\boldsymbol{y},\boldsymbol{\eta}) - 1 + O_p(N^{-1/2-r}), \tag{3.9}$$

where $A_B(\boldsymbol{y},\boldsymbol{\eta}) - 1 = O_p(N^{-r})$ *for* $A_B(\boldsymbol{y},\boldsymbol{\eta})$ *defined by*

$$A_B(\boldsymbol{y},\boldsymbol{\eta}) = [1 + \Delta_v(\boldsymbol{y})/\nu_0]^{1/2}.$$

Proof. It follows from (2.7) and condition (A5) that

$$\begin{aligned}\sum_{i=1}^{K} w_i\{\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}}) - \overline{\widehat{\xi}}_w(\boldsymbol{y},\widehat{\boldsymbol{\eta}})\}^2 &= \sum_{i=1}^{K} w_i\{\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta}) - \overline{\widehat{\xi}}_w(\boldsymbol{y},\boldsymbol{\eta})\}^2 + O_p(N^{-1/2})\\ &= \sum_{i=1}^{K} w_i E[\{\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta}) - \overline{\widehat{\xi}}_w(\boldsymbol{y},\boldsymbol{\eta})\}^2] + O_p(N^{-1/2}),\end{aligned}$$

which shows (3.8). Using the approximation (3.8), we can estimate $a_B(\boldsymbol{y}) - 1$ as

$$\begin{aligned}a_B(\boldsymbol{y}) - 1 =& \frac{\{a_B(\boldsymbol{y})\}^2 - 1}{a_B(\boldsymbol{y}) + 1} = \frac{\Delta_v(\boldsymbol{y})/\nu_0}{1 + [1 + \Delta_v(\boldsymbol{y})/\nu_0]^{1/2}} + O_p(N^{-r-1/2})\\ =& \frac{\Delta_v(\boldsymbol{y})/\nu_0\{1 - [1 + \Delta_v(\boldsymbol{y})/\nu_0]^{1/2}\}}{1 - [1 + \Delta_v(\boldsymbol{y})/\nu_0]} + O_p(N^{-r-1/2})\\ =& [1 + \Delta_v(\boldsymbol{y})/\nu_0]^{1/2} - 1 + O_p(N^{-r-1/2}).\end{aligned}$$

This shows that $a_B(\boldsymbol{y}) - 1 = O_p(N^{-r})$ as well as (3.9). ■

Using Lemma 3.1, we can see that $I_2(\boldsymbol{\eta})$ is estimated as $I_2(\boldsymbol{\eta}) = E[\{h(\boldsymbol{y},\boldsymbol{\eta})\}^2]$, where

$$\begin{aligned}h(\boldsymbol{y},\boldsymbol{\eta}) =& \Big[A_B(\boldsymbol{y},\boldsymbol{\eta}) - 1 + O_p(N^{-1/2-r})\Big]\Big[\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta}) - \sum_{j=1}^{K} w_j E[\widehat{\xi}_j(\boldsymbol{y},\boldsymbol{\eta})]\\ &- \sum_{j=1}^{K} w_j\big\{\widehat{\xi}_j(\boldsymbol{y},\boldsymbol{\eta}) - E[\widehat{\xi}_j(\boldsymbol{y},\boldsymbol{\eta})]\big\} + (\frac{\partial\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}})^t(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})\\ &- \sum_{j=1}^{K} w_j(\frac{\partial\widehat{\xi}_j(\boldsymbol{y},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}})^t(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) + O_p(N^{-1})\Big] + \Delta_m(\boldsymbol{y}).\end{aligned} \tag{3.10}$$

Since $\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta}) = E[\xi_i|\boldsymbol{y}]$, it is also noted that $I_3(\boldsymbol{\eta})$ is evaluated as

$$\begin{aligned}I_3(\boldsymbol{\eta}) =& E[\big\{\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}}) - \widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta})\big\}\big\{(a_B(\boldsymbol{y}) - 1)(\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}}) - \overline{\widehat{\xi}}_w(\boldsymbol{y},\widehat{\boldsymbol{\eta}})) + \Delta_m(\boldsymbol{y})\big\}]\\ =& E[(\frac{\partial\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}})^t(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})h(\boldsymbol{y},\boldsymbol{\eta})] + O(N^{-1-r}) + o(N^{-1}).\end{aligned} \tag{3.11}$$

We handle asymptotic properties of $I_2(\boldsymbol{\eta})$ and $I_3(\boldsymbol{\eta})$ for the cases of $r=0$, $r=1/2$ and $r=1$ separately.

(Scenario 1: $r=0$) In this case, it can be seen that $I_2(\boldsymbol{\eta})=O(1)$, and $I_2(\boldsymbol{\eta})$ is approximated as

$$I_2(\boldsymbol{\eta})=E\Big[\big\{A_B(\boldsymbol{y},\boldsymbol{\eta})-1\big\}^2\big\{\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta})-\sum_{j=1}^K w_j E[\widehat{\xi}_j(\boldsymbol{y},\boldsymbol{\eta})]\big\}^2\Big]+O(N^{-1/2}).$$

Since $I_3(\boldsymbol{\eta})=O(N^{-1/2})$, we get the following proposition.

Proposition 3.2 *Assume conditions* (A1)-(A5). *Then, CEB subject to the mean-variance constraints* (C1) *and* (C2) *with $r=0$ is given by* (3.4), *and*

$$\begin{aligned}&\lim_{N\to\infty}\{MSE(\boldsymbol{\eta},\delta_i^{CEB}(\boldsymbol{y},\widehat{\boldsymbol{\eta}}))-MSE(\boldsymbol{\eta},\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}}))\}\\&\quad=\lim_{N\to\infty}E\Big[\big\{A_B(\boldsymbol{y},\boldsymbol{\eta})-1\big\}^2\big\{\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta})-\sum_{j=1}^K w_j E[\widehat{\xi}_j(\boldsymbol{y},\boldsymbol{\eta})]\big\}^2\Big].\end{aligned}$$

This proposition means that in the mean-variance constraints with $r=0$, the difference in MSE between $\delta_i^{CEB}(\boldsymbol{y},\widehat{\boldsymbol{\eta}})$ and EB $\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}})$ appears in the first-order term.

(Scenario 2: $r=1/2$) In this case, it can be seen from (3.10) that $I_2(\boldsymbol{\eta})=O(N^{-1})$, and $I_2(\boldsymbol{\eta})$ is approximated as $I_2(\boldsymbol{\eta})=E[\{A_B(\boldsymbol{y},\boldsymbol{\eta})-1\}^2\{\overline{h}(\boldsymbol{y},\boldsymbol{\eta})\}^2]+o(N^{-1})$, where

$$\overline{h}(\boldsymbol{y},\boldsymbol{\eta})=\Big[A_B(\boldsymbol{y},\boldsymbol{\eta})-1\Big]\Big[\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta})-\sum_{j=1}^K w_j E[\widehat{\xi}_j(\boldsymbol{y},\boldsymbol{\eta})]\Big]+\Delta_m(\boldsymbol{y}). \tag{3.12}$$

Similarly,

$$I_3(\boldsymbol{\eta})=E[(\frac{\partial\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}})^t(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})\overline{h}(\boldsymbol{y},\boldsymbol{\eta})]+o(N^{-1}).$$

Proposition 3.3 *Assume conditions* (A1)-(A5). *Then, CEB subject to the mean-variance constraints* (C1) *and* (C2) *with $r=1/2$ is given by* (3.4), *and*

$$\begin{aligned}MSE(\boldsymbol{\eta},\delta_i^{CEB}(\boldsymbol{y},\widehat{\boldsymbol{\eta}}))=&MSE(\boldsymbol{\eta},\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}}))+E[\{A_B(\boldsymbol{y},\boldsymbol{\eta})-1\}^2\{\overline{h}(\boldsymbol{y},\boldsymbol{\eta})\}^2]\\&+2E[(\frac{\partial\widehat{\xi}_i(\boldsymbol{y},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}})^t(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})\overline{h}(\boldsymbol{y},\boldsymbol{\eta})]+o(N^{-1}).\end{aligned}$$

This proposition means that in the mean-variance constraints with $r=1/2$, the difference in MSE between $\delta_i^{CEB}(\boldsymbol{y},\widehat{\boldsymbol{\eta}})$ and EB $\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}})$ appears in the second-order term.

(Scenario 3: $r=1$) In this case, from (3.10), we get the same approximation as in Proposition 3.1.

Proposition 3.4 *Assume conditions* (A1)-(A5)*. Then, CEB subject to the mean-variance constraints* (C1) *and* (C2) *with $r = 1$ is given by* (3.4)*, and*

$$MSE(\boldsymbol{\eta}, \delta_i^{CEB}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}})) = MSE(\boldsymbol{\eta}, \widehat{\xi}_i(\boldsymbol{y}, \widehat{\boldsymbol{\eta}})) + E[\{\Delta_m(\boldsymbol{y})\}^2] \\ + 2E\Big[\Big(\frac{\partial \widehat{\xi}_i(\boldsymbol{y}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\Big)^t (\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})\Delta_m(\boldsymbol{y})\Big] + o(N^{-1}).$$

3.3 Estimation of MSE of CEB

Finally, we provide a second-order unbiased estimator of MSE for CEB subject to the mean-variance constraints (C1) and (C2). Since the second-order unbiased estimator of $MSE(\boldsymbol{\eta}, \widehat{\xi}_i(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}))$ is given in (2.10), second-order unbiased estimators can be derived based on the Taylor series expansion from Propositions 3.1, 3.3 and 3.4. However, it may be hard to derive a second-order unbiased estimator for $r = 0$. Instead of them, we here suggest a procedure using the parametric bootstrap method for all $r \geq 0$. To this end, we assume the following condition:

(A6) $\Delta_v(\boldsymbol{y}) - E[\Delta_v(\boldsymbol{y})] = O_p(N^{-r-1/2})$, $E[\{\widehat{\xi}_i(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}) - \widehat{\xi}_i(\boldsymbol{y}, \boldsymbol{\eta})\}\widehat{\xi}_i(\boldsymbol{y}, \boldsymbol{\eta})] = O(N^{-1})$ and $E[\widehat{\xi}_i(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}) - \widehat{\xi}_i(\boldsymbol{y}, \boldsymbol{\eta})] = O(N^{-1})$.

As an estimator of $I_3(\boldsymbol{\eta})$ in (3.6), we consider $I_3^*(\boldsymbol{y})$ given by

$$I_3^*(\boldsymbol{y}) = E^*[\{\widehat{\xi}_i(\boldsymbol{y}^*, \widehat{\boldsymbol{\eta}}^*) - \widehat{\xi}_i(\boldsymbol{y}^*, \widehat{\boldsymbol{\eta}})\}\{(a_B(\boldsymbol{y}^*) - 1)(\widehat{\xi}_i(\boldsymbol{y}^*, \widehat{\boldsymbol{\eta}}^*) - \overline{\widehat{\xi}}_w(\boldsymbol{y}^*, \widehat{\boldsymbol{\eta}}^*)) + \Delta_m(\boldsymbol{y}^*)\}].$$

An exact unbiased estimator of $I_2(\boldsymbol{\eta})$ in (3.6) is $\widehat{I}_2(\boldsymbol{y})$ given by

$$\widehat{I}_2(\boldsymbol{y}) = \{(a_B(\boldsymbol{y}) - 1)(\widehat{\xi}_i(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}) - \overline{\widehat{\xi}}_w(\boldsymbol{y}, \widehat{\boldsymbol{\eta}})) + \Delta_m(\boldsymbol{y})\}^2.$$

Theorem 3.1 *Assume conditions* (A1)-(A6)*. Then, CEB subject to the mean-variance constraints* (C1) *and* (C2) *with $r \geq 0$ is given by* (3.4)*, and a second-order unbiased estimator of $MSE(\boldsymbol{\eta}, \delta_i^{CEB}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}))$ based on the parametric bootstrap method is*

$$mse(\boldsymbol{y}, \delta_i^{CEB}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}})) = mse^*(\boldsymbol{y}, \widehat{\xi}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}})) + \widehat{I}_2(\boldsymbol{y}) + 2I_3^*(\boldsymbol{y}), \tag{3.13}$$

namely, $E[mse(\boldsymbol{y}, \delta_i^{CEB}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}))] = MSE(\boldsymbol{\eta}, \delta_i^{CEB}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}})) + o(N^{-1})$, where $mse^*(\boldsymbol{y}, \widehat{\xi}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}))$ is given in (2.12)*.*

Proof. Theorem 2.3 implies that $E[mse^*(\boldsymbol{y}, \widehat{\xi}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}))] = I_1(\boldsymbol{\eta}) + o(N^{-1})$. Clearly, $E[\widehat{I}_2(\boldsymbol{y})] = I_2(\boldsymbol{\eta})$. If $I_3(\boldsymbol{\eta}) = O(N^{-1})$, we can verify that $E[I_3^*(\boldsymbol{y})] = I_3(\boldsymbol{\eta}) + o(N^{-1})$. Thus, we need to show that $I_3(\boldsymbol{\eta}) = O(N^{-1})$. It follows from (3.10) and (3.11) that

$$I_3(\boldsymbol{\eta}) = E[\{\widehat{\xi}_i(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}) - \widehat{\xi}_i(\boldsymbol{y}, \boldsymbol{\eta})\}\{(a_B(\boldsymbol{y}) - 1)(\widehat{\xi}_i(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}) - \overline{\widehat{\xi}}_w(\boldsymbol{y}, \widehat{\boldsymbol{\eta}})) + \Delta_m(\boldsymbol{y})\}] \\ = E[\{\widehat{\xi}_i(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}) - \widehat{\xi}_i(\boldsymbol{y}, \boldsymbol{\eta})\}\{A_B(\boldsymbol{y}, \boldsymbol{\eta}) - 1\}\{\widehat{\xi}_i(\boldsymbol{y}, \boldsymbol{\eta}) - \sum_{j=1}^K w_j E[\widehat{\xi}_j(\boldsymbol{y}, \boldsymbol{\eta})]\}] + O(N^{-1}).$$

As seen from the proof of Lemma 3.1, it is noted that

$$A_B(\boldsymbol{y}, \boldsymbol{\eta}) - 1 = \frac{\Delta_v(\boldsymbol{y})/\nu_0}{1 + [1 + \Delta_v(\boldsymbol{y})/\nu_0]^{1/2}},$$

so that

$$A_B(\boldsymbol{y}, \boldsymbol{\eta}) - 1 - \frac{E[\Delta_v(\boldsymbol{y})]/\nu_0}{1 + [1 + E[\Delta_v(\boldsymbol{y})]/\nu_0]^{1/2}} = O_p(N^{-1/2}),$$

from condition (A6). Hence,

$$\begin{aligned} I_3(\boldsymbol{\eta}) =& \frac{E[\Delta_v(\boldsymbol{y})]/\nu_0}{1 + [1 + E[\Delta_v(\boldsymbol{y})]/\nu_0]^{1/2}} \Big\{ E[\{\widehat{\xi}_i(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}) - \widehat{\xi}_i(\boldsymbol{y}, \boldsymbol{\eta})\}\widehat{\xi}_i(\boldsymbol{y}, \boldsymbol{\eta})] \\ & - E[\widehat{\xi}_i(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}) - \widehat{\xi}_i(\boldsymbol{y}, \boldsymbol{\eta})] \sum_{j=1}^{K} w_j E[\widehat{\xi}_j(\boldsymbol{y}, \boldsymbol{\eta})]] \Big\} + O(N^{-1}), \end{aligned}$$

which is of order $O(N^{-1})$ under condition (A6). Therefore, we get the result of the theorem. ■

4 Applications to NEF-QVF

4.1 CEB in NEF-QVF and previous results on MSE

We now apply the results in the previous section to natural exponential families with quadratic variance functions (NEF-QVF). The small-area estimation based on NEF-QVF was studied by Ghosh and Maiti (2004), who treated the unit level model with individual observation having NEF-QVF. In this section, we handle an area level model with a survey estimate from each small area having NEF-QVF, and try to apply the results in the previous section to the benchmarking problem in NEF-QVF using the results of Ghosh and Maiti (2004).

Let $y_1, \ldots, y_K$ 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:

$$\begin{aligned} y_i|\theta_i \sim & f(y_i|\theta_i) = \exp[n_i(\theta_i y_i - \psi(\theta_i)) + c(y_i, n_i)], \\ \theta_i|\nu, m_i \sim & \pi(\theta_i|\nu, m_i) = \exp[\nu(m_i\theta_i - \psi(\theta_i))]C(\nu, m_i), \end{aligned} \tag{4.1}$$

where n_i is a known scalar and ν is an unknown scalar. Let $\boldsymbol{y} = (y_1, \ldots, y_K)^t$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^t$. 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 μ_i by

$$\mu_i = E[y_i|\theta_i] = \psi'(\theta_i).$$

Then, we assume that $\psi''(\theta_i) = Q_i(\mu_i)$, namely,

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

where $Q_i(x) = v_{0,i} + v_{1,i}x + v_{2,i}x^2$ for known constants $v_{0,i}$, $v_{1,i}$ and $v_{2,i}$ which are not simultaneously zero. This is the natural exponential family with the quadratic variance function (NEF-QVF) treated by Morris (1982, 83). The binomial distribution $\mathcal{B}in(n_i, \mu_i/n_i)$ for $\mu_i = n_i p_i$ corresponds to $v_{0,i} = 0$, $v_{1,i} = n_i$ and $v_{2,i} = -1$. The Poisson distribution $\mathcal{P}o(\mu_i)$ for $\mu_i = n_i\lambda_i$ corresponds to $v_{0,i} = v_{2,i} = 0$ and $v_{1,i} = n_i$. For the normal distribution $\mathcal{N}(\mu_i, \sigma^2/n_i)$ with known variance σ^2, we have $v_{0,i} = \sigma^2$ and $v_{1,i} = v_{2,i} = 0$. Similarly, the mean and variance of the prior distribution are given by

$$E[\mu_i|m_i, \nu] = m_i, \quad Var(\mu_i|m_i, \nu) = \frac{Q_i(m_i)}{\nu - v_{2,i}}. \tag{4.2}$$

In this section, we consider the canonical link, namely

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

The unknown parameters $\boldsymbol{\eta}$ in the previous sections 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

$$\begin{aligned} \pi(\theta_i|y_i, \nu, m_i) =& \exp[(n_i + \nu)(\widehat{\mu}_i\theta_i - \psi(\theta_i))] C(n_i + \nu, \widehat{\mu}_i), \\ f_\pi(y_i|\nu, m_i) =& \frac{C(\nu, m_i)}{C(n_i + \nu, \widehat{\mu}_i)} \exp[c(y_i, n_i)], \end{aligned} \tag{4.3}$$

where $\widehat{\mu}_i$ is the conditional expectation of μ_i given y_i, namely, $\widehat{\mu}_i = E[\mu_i|y_i, \boldsymbol{\eta}]$, given by

$$\widehat{\mu}_i = \widehat{\mu}_i(y_i, \boldsymbol{\eta}) = \frac{n_i y_i + \nu m_i}{n_i + \nu}.$$

In the binomial and Poisson distributions, the parameters one want to estimate are $p_i = \mu_i/n_i$ and $\lambda_i = \mu_i/n_i$, respectively. Thus, in this section, we shall estimate

$$\xi_i = q_i \mu_i, \tag{4.4}$$

for known positive constant q_i. When ν and m_i are known, the Bayes estimator of ξ_i in the Bayesian context is

$$\widehat{\xi}_i(y_i, \boldsymbol{\eta}) = q_i \widehat{\mu}_i(y_i, \boldsymbol{\eta}) = q_i \frac{n_i y_i + \nu m_i}{n_i + \nu}. \tag{4.5}$$

As shown by Ghosh and Maiti (2004),

$$\begin{aligned} E[y_i] =& E[\psi'(\theta_i)] = m_i, \\ Var(y_i) =& Var(E[y_i|\theta_i]) + E[Var(y_i|\theta_i)] = Var(\mu_i) + E[Q_i(\mu_i)/n_i] = Q_i(m_i)\phi_i, \\ Cov(y_i, \mu_i) =& E[Cov(y_i, \mu_i)|\theta_i] + Cov(E[y_i|\theta_i], \mu_i) = Q_i(m_i)/(\nu - v_{2,i}), \end{aligned}$$

for $\phi_i = (1+\nu/n_i)/(\nu - v_{2,i})$. Using these observations, Ghosh and Maiti (2004) showed that the Bayes estimator $\widehat{\xi}_i$ given in (4.5) is the best linear unbiased predictor (BLUP) of ξ_i in terms of MSE.

Following Godambe and Thompson (1989), Ghosh and Maiti (2004) proposed the estimators of $\boldsymbol{\beta}$ and ν through the estimating equations. 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

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

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, the optimal estimating equations given by $\sum_{i=1}^K \boldsymbol{D}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{g}_i = \mathbf{0}$ are written as

$$
\begin{aligned}
&\sum_{i=1}^K \frac{1}{|\boldsymbol{\Sigma}_i|}\Big[\{\mu_{4i} - \mu_{2i}^2 - \mu_{3i}\phi_i Q_i'(m_i)\}g_{1i} + \{\mu_{2i}\phi_i Q_i'(m_i) - \mu_{3i}\}g_{2i}\Big] Q_i(m_i)\boldsymbol{x}_i = \mathbf{0}, \\
&\sum_{i=1}^K \frac{1}{|\boldsymbol{\Sigma}_i|}\{\mu_{2i}g_{2i} - \mu_{3i}g_{1i}\} Q_i(m_i)(1+v_{2,i}/n_i)(\nu - v_{2,i})^{-2} = 0.
\end{aligned}
\tag{4.6}
$$

Solving the equations on $\boldsymbol{\eta}^t = (\boldsymbol{\beta}^t, \hat{\nu})$ simultaneously, we obtain estimator $\widehat{\boldsymbol{\eta}}^t = (\widehat{\boldsymbol{\beta}}^t, \hat{\nu})$. Letting $\hat{m}_i = \psi'(\boldsymbol{x}_i^t \widehat{\boldsymbol{\beta}})$ and substituting $\hat{m}_i$ and $\hat{\nu}$ into (4.5), we get EB

$$
\widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}}) = q_i \frac{n_i y_i + \hat{\nu}\hat{m}_i}{n_i + \hat{\nu}}. \tag{4.7}
$$

As shown in the appendix of Ghosh and Maiti (2004),

$$
\begin{aligned}
&\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta} = \boldsymbol{U}_K^{-1}\boldsymbol{s}_k(\boldsymbol{\eta}) + o_p(K^{-1/2}), \\
&E[(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t] = \boldsymbol{U}_K^{-1} + o(K^{-1}),\ E[\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}] = O(K^{-1}),
\end{aligned}
\tag{4.8}
$$

where $\boldsymbol{s}_k(\boldsymbol{\eta}) = \sum_{i=1}^K \boldsymbol{D}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{g}_i$ and $\boldsymbol{U}_K = \mathbf{Cov}\,(\boldsymbol{s}_K(\boldsymbol{\eta})) = \sum_{i=1}^K \boldsymbol{D}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{D}_i$, which is of order $O(K)$. These show condition (A1). Clearly, condition (A2) is satisfied. Since $Var(\mu_i|y_i) = Q_i(\mu_i)/(n_i + \nu - v_{2,i})$, it is observed that

$$
g_1(\boldsymbol{\eta}) = \frac{E[Q_i(\mu_i)]}{n_i + \nu - v_{2,i}} = q_i^2 \frac{\nu}{(n_i+\nu)(\nu - v_{2,i})} Q_i(m_i),
$$

so that condition (A3) is satisfied. Thus, Theorems 2.1, 2.2 and 2.3 hold, where a detailed expression of $g_2(\boldsymbol{\eta})$ is given in Ghosh and Maiti (2004).

Finally, we give the exact moments of μ_{ri} following Ghosh and Maiti (2004). The proof is given in the Appendix.

Proposition 4.1 *The moments* $\mu_{ri} = E[(y_i - m_i)^r]$, $r = 2, 3, 4$, *are written as*

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

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

for $d_i = v_{2,i}/n_i$. *These are identical to the same quantities given in Theorem 1 of Ghosh and Maiti (2004), where* δ_i, ξ_i *and* ν_i *in their paper correspond to* n_i^{-1}, n_i^{-2} *and* n_i^{-3}. *Further, according to Morris (1983) and Ghosh and Maiti (2004), it is seen that* $E[(\mu_i - m_i)^2] = Q_i(m_i)/(\nu - v_{2,i})$, $E[(\mu_i - m_i)^3] = 2Q_i(m_i)Q_i'(m_i)/\{(\nu - v_{2,i})(\nu - 2v_{2,i})\}$ *and*

$$E[(\mu_i - m_i)^4] = \frac{3Q_i(m_i)[(\nu - 2v_{2,i})Q_i(m_i) + 2\{Q_i'(m_i)\}^2]}{(\nu - v_{2,i})(\nu - 2v_{2,i})(\nu - 3v_{2,i})}.$$

4.2 MSE and estimation of MSE for CEB

We now apply the results in Section 3 to the NEF-QVF distributions. As specific constraints of (C1) and (C2), we here consider the following for estimator $\delta_i(\boldsymbol{y})$ of $\xi_i = q_i\mu_i$:

(C1′) $\sum_{i=1}^K w_i\delta_i(\boldsymbol{y}) = \sum_{i=1}^K w_i q_i y_i$,

(C2′) $\sum_{i=1}^K w_i\{\delta_i(\boldsymbol{y}) - \overline{\delta}_w(\boldsymbol{y})\}^2 = \sum_{i=1}^K w_i\{\widehat{\xi}_i(y_i, \widehat{\boldsymbol{\eta}}) - \overline{\widehat{\xi}}_w(\boldsymbol{y}, \widehat{\boldsymbol{\eta}})\}^2 + \Delta_v^\dagger(\boldsymbol{y})$, where

$$\Delta_v^\dagger(\boldsymbol{y}) = K^{-r}\sum_{i=1}^K w_i(1 - w_i)q_i^2\frac{Q_i(\widehat{\mu}_i(y_i, \widehat{\boldsymbol{\eta}}))}{n_i + \hat{\nu} - v_{2,i}}, \tag{4.9}$$

where $\overline{\delta}_w(\boldsymbol{y}) = \sum_{i=1}^K w_i\delta_i(\boldsymbol{y})$ and $\overline{\widehat{\xi}}_w(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}) = \sum_{i=1}^K w_i\widehat{\xi}_i(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}) = \sum_{i=1}^K w_i q_i\widehat{\mu}_i(\boldsymbol{y}, \widehat{\boldsymbol{\eta}})$.

It is noted that $\Delta_m(\boldsymbol{y})$ in (C1) is expressed as

$$\Delta_m^\dagger(\boldsymbol{y}) = \sum_{i=1}^K \frac{w_i q_i}{n_i + \hat{\nu}}\hat{\nu}(y_i - \hat{m}_i), \tag{4.10}$$

which can be seen to be of order $O_p(K^{-1/2})$. Also, it is noted that $N^r\Delta_v(\boldsymbol{y})$ is an estimator of the second term in the r.h.s. of (3.1). In fact,

$$\begin{aligned}\sum_{i=1}^K w_i Var(\xi_i - \overline{\widehat{\xi}}_w|\boldsymbol{y}) =&\sum_{i=1}^K w_i\Big\{q_i^2(1 - 2w_i)Var(\mu_i|y_i) + \sum_{j=1}^K w_j^2 q_j^2 Var(\mu_j|y_j)\Big\}\\ =&\sum_{i=1}^K (w_i q_i^2 - 2w_i^2 q_i^2 + w_i^2 q_i^2)Var(\mu_i|y_i)\\ =&\sum_{i=1}^K w_i(1 - w_i)q_i^2 Q_i(\mu)/(n_i + \nu - v_{2,i}),\end{aligned}$$

which is of order $O_p(1)$. Under constraints (C1′) and (C2′), it follows from (3.4) and (3.5) that CEB is

$$\delta_i^{CEB}(\boldsymbol{y},\widehat{\boldsymbol{\eta}}) = \widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}}) + \{a_B(\boldsymbol{y})-1\}\{\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}}) - \overline{\widehat{\xi}}_w(\boldsymbol{y},\widehat{\boldsymbol{\eta}})\} + \Delta_m^\dagger(\boldsymbol{y}), \tag{4.11}$$

where

$$\{a_B(\boldsymbol{y})\}^2 = 1 + \frac{\Delta_v^\dagger(\boldsymbol{y})}{\sum_{i=1}^K w_i\{\widehat{\xi}_i(\boldsymbol{y},\widehat{\boldsymbol{\eta}}) - \overline{\widehat{\xi}}_w(\boldsymbol{y},\widehat{\boldsymbol{\eta}})\}^2}, \tag{4.12}$$

for $\Delta_m^\dagger(\boldsymbol{y})$ and $\Delta_v^\dagger(\boldsymbol{y})$ given in (4.10) and (4.9).

Lemma 4.1 *Assume that* $\max_i\{n_i\}$, $\max_i\{m_i\}$ *and* $\max_i\{q_i\}$ *are bounded, and that* $\sum_{i=1}^K w_i = 1$ *and* $\sum_{i=1}^K w_i^2 = O(K^{-1})$. *Then conditions* (A4), (A5) *and* (A6) *are satisfied.*

Proof. We first check conditions (A4) and (A5). For notational simplicity, let $\widehat{\mu}_i = \widehat{\mu}_i(y_i,\boldsymbol{\eta})$ and $\widehat{\xi}_i = \widehat{\xi}_i(y_i,\boldsymbol{\eta})$. It is noted that $\widehat{\mu}_i - m_i = n_i(y_i-m_i)/(n_i+\nu)$, so that

$$\sum_{i=1}^K w_i\{\widehat{\xi}_i - E[\widehat{\xi}_i]\} = \sum_{i=1}^K w_i q_i \frac{n_i}{n_i+\nu}(y_i-m),$$

which is of order $O_p(K^{-1/2})$. Thus, condition (A4) is satisfied. For (A5), it is noted that

$$\begin{aligned}
\sum_{i=1}^K w_i(\widehat{\xi}_i - \overline{\widehat{\xi}}_w)^2 =& \sum_{i=1}^K w_i\Big\{q_i(\widehat{\mu}_i - m_i) - \sum_{j=1}^K w_jq_j(\widehat{\mu}_j-m_j) + (q_im_i - \sum_{j=1}^K w_jq_jm_j)\Big\}^2\\
=& \sum_{i=1}^K w_iq_i^2(\widehat{\mu}_i-m_i)^2 - \{\sum_{i=1}^K w_iq_i(\widehat{\mu}_i-m_i)\}^2 + \{\sum_{i=1}^K\big(q_im_i - \sum_{j=1}^K w_jq_jm_j\big)\}^2\\
&+ 2\sum_{i=1}^K w_iq_i(\widehat{\mu}_i-m_i)(q_im_i - \sum_{j=1}^K w_jq_jm_j)\\
&- 2\{\sum_{i=1}^K w_i(q_im_i - \sum_{j=1}^K w_jq_jm_j)\}\{\sum_{i=1}^K w_iq_i(\widehat{\mu}_i-m_i)\}.
\end{aligned}$$

so that we have

$$\begin{aligned}
\sum_{i=1}^K w_i(\widehat{\xi}_i - \overline{\widehat{\xi}}_w)^2 - \sum_{i=1}^K w_iE[(\widehat{\xi}_i - \overline{\widehat{\xi}}_w)^2] =& \sum_{i=1}^K w_iq_i^2\{(\widehat{\mu}_i-m_i)^2 - E[(\widehat{\mu}_i-m_i)^2]\}\\
&- \Big\{\{\sum_{i=1}^K w_iq_i(\widehat{\mu}_i-m_i)\}^2 - E[\{\sum_{i=1}^K w_iq_i(\widehat{\mu}_i-m_i)\}^2]\Big\}\\
&+ 2\sum_{i=1}^K w_iq_i(\widehat{\mu}_i-m_i)(q_im_i - \sum_{j=1}^K w_jq_jm_j)\\
&- 2\{\sum_{i=1}^K w_i(q_im_i - \sum_{j=1}^K w_jq_jm_j)\}\{\sum_{i=1}^K w_iq_i(\widehat{\mu}_i-m_i)\}\\
=& I_1 - I_2 + 2I_3 - 2I_4. \qquad \text{(say)}
\end{aligned}$$

Since $\sum_{i=1}^K w_i = 1$ and $\sum_{i=1}^K w_i^2 = O(K^{-1})$, it can be seen that $I_3 = O_p(K^{-1/2})$ and $I_4 = O_p(K^{-1/2})$. For I_1, it is noted that

$$E[\{\sum_{i=1}^K w_i q_i \frac{n_i^2}{(n_i+\nu)^2}\{(y_i-m_i)^2 - E[(y_i-m_i)^2]\}\}^2] = \sum_{i=1}^K w_i^2 q_i^2 \frac{n_i^4}{(n_i+\nu)^4}(\mu_{4i}-\mu_{2i}^2),$$

which is of $O(K^{-1})$, where μ_{4i} and μ_{2i} are given in Proposition 4.1. Thus, $I_1 = O_p(K^{-1/2})$. Further, I_2 is rewritten as $I_2 = I_{21} + 2I_{22}$, where

$$I_{21} = \sum_{i=1}^K w_i^2 q_i^2 \frac{n_i^2}{(n_i+\nu)^2}\{(y_i-m_i)^2 - E[(y_i-m_i)^2]\},$$

$$I_{22} = \sum_{i=1}^K \sum_{j>i} \frac{w_i q_i n_i}{n_i+\nu}\frac{w_j q_j n_j}{n_j+\nu}(y_i-m_i)(y_j-m_j).$$

Similarly to I_1, it can be seen that $I_{21} = O_p(K^{-1/2})$. For I_{22}, it is noted that for $c_i = w_i q_i n_i/(n_i+\nu)$,

$$\begin{aligned}
E[\{I_{22}\}^2] =& \sum_i \sum_{j>i} \sum_k \sum_{\ell>k} E[c_i(y_i-m_i)c_j(y_j-m_j)c_k(y_k-m_k)c_\ell(y_\ell-m_\ell)] \\
=& \sum_i \sum_{j>i} \sum_{\ell>i} E[c_i^2(y_i-m_i)^2]E[c_j(y_j-m_j)c_\ell(y_\ell-m_\ell)] \\
&+ 2\sum_i \sum_{j>i} \sum_{k>i} \sum_{\ell>k} E[c_i(y_i-m_i)c_j(y_j-m_j)c_k(y_k-m_k)c_\ell(y_\ell-m_\ell)] \\
=& \sum_i \sum_{j>i} \sum_{\ell>i} E[c_i^2(y_i-m_i)^2]E[c_j(y_j-m_j)c_\ell(y_\ell-m_\ell)] \\
=& \sum_i \sum_{j>i} E[c_i^2(y_i-m_i)^2]E[c_j^2(y_j-m_j)^2] \\
&+ \sum_i \sum_{j>i} \sum_{\ell>i,\ell\neq j} E[c_i^2(y_i-m_i)^2]E[c_j(y_j-m_j)]E[c_\ell(y_\ell-m_\ell)] \\
=& 2^{-1}\sum_i \sum_{j\neq i} E[c_i^2(y_i-m_i)^2]E[c_j^2(y_j-m_j)^2] \\
\leq& 2^{-1}\Big\{\sum_{i=1}^K \frac{w_i^2 q_i^2 n_i^2}{(n_i+\nu)^2} Q_i(m_i)\phi_i\Big\}^2,
\end{aligned}$$

which is of order $O(K^{-1})$. Thus, these observations show that condition (A5) is satisfied.

Finally, we shall check condition (A6). Note that

$$\begin{aligned}
\widehat{\mu}_i(y_i, \widehat{\boldsymbol{\eta}}) - \widehat{\mu}_i(y_i, \boldsymbol{\eta}) =& \frac{n_i y_i + \hat{\nu}\hat{m}_i}{n_i+\hat{\nu}} - \frac{n_i y_i + \nu m_i}{n_i+\nu} \\
=& -\frac{n_i(\hat{\nu}-\nu)(y_i-m_i)}{(n_i+\hat{\nu})(n_i+\nu)} + n_i m_i(\nu-\hat{\nu})\frac{(\hat{\nu}\hat{m}_i - \nu m_i)n_i + (\hat{m}_i - m_i)\hat{\nu}\nu}{(n_i+\hat{\nu})(n_i+\nu)} \\
=& -\frac{n_i(\hat{\nu}-\nu)(y_i-m_i)}{(n_i+\hat{\nu})(n_i+\nu)} + O_p(K^{-1}), \qquad (4.13)
\end{aligned}$$

where the approximation in the last equality follows from (4.8). Since $\Delta_v^\dagger(\boldsymbol{y})$ given in (4.9) can be approximated as $\Delta_v^\dagger(\boldsymbol{y}) = K^{-r}\sum_{i=1}^K w_i(1-w_i)q_i^2 Q_i(\widehat{\mu}_i(y_i,\boldsymbol{\eta}))/(n_i+\nu-v_{2,i}) + O_p(K^{-r-1/2})$, it is observed that

$$\begin{aligned}\Delta_v(\boldsymbol{y}) - E[\Delta_v(\boldsymbol{y})] =& K^{-r}\sum_{i=1}^K \frac{w_i(1-w_i)q_i^2}{n_i+\nu-v_{2,i}}\Big\{(v_1+2v_{2,i}m_i)\frac{n_i(y_i-m_i)}{n_i+\nu} \\ &+ v_{2,i}\frac{n_i^2}{(n_i+\nu)^2}\{(y_i-m_i)^2 - E[(y_i-m_i)^2]\}\Big\} + O_p(K^{-r-1/2}),\end{aligned}$$

which can be verified to be of order $O_p(K^{-r-1/2})$. It follows from (4.13) that

$$E[\widehat{\xi}_i(y_i,\widehat{\boldsymbol{\eta}}) - \widehat{\xi}_i(y_i,\boldsymbol{\eta})] = -\frac{q_i n_i E[(\hat{\nu}-\nu)(y_i-m_i)]}{(n_i+\nu)^2} + O(K^{-1}).$$

Similarly,

$$\begin{aligned}&E[\{\widehat{\xi}_i(y_i,\widehat{\boldsymbol{\eta}}) - \widehat{\xi}_i(y_i,\boldsymbol{\eta})\}\widehat{\xi}_i(y_i,\boldsymbol{\eta})] \\ &\quad = -\frac{q_i^2 n_i^2 E[(\hat{\nu}-\nu)(y_i-m_i)^2]}{(n_i+\nu)^3} - \frac{q_i^2 n_i m_i E[(\hat{\nu}-\nu)(y_i-m_i)]}{(n_i+\nu)^2} + O(K^{-1}).\end{aligned}$$

Thus, it is sufficient to show that $E[(\hat{\nu}-\nu)(y_i-m_i)] = O(K^{-1})$ and $E[(\hat{\nu}-\nu)(y_i-m_i)^2] = O(K^{-1})$. Recall the facts given in (4.8). Note that $\boldsymbol{U}_K^{-1} = O(K^{-1})$ and

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\nu}-\nu \end{pmatrix} = \boldsymbol{U}_K^{-1}\sum_{j=1}^K \boldsymbol{D}_j^t\boldsymbol{\Sigma}_j^{-1}\begin{pmatrix} y_j - m_j \\ (y_j-m_j)^2 - \phi_j Q_i(m_j)\end{pmatrix}.$$

Let $\boldsymbol{e}_0 = (0,\ldots,0,1)^t$ be a vector with the last component one and the others zeros. Since $E[\{(y_j-m_j)^2 - \phi_j Q_i(m_j)\}(y_i-m_i)] = 0$ for $i \neq j$, it is seen that

$$\begin{aligned}E[(\hat{\nu}-\nu)(y_i-m_i)] =& \boldsymbol{e}_0^t\boldsymbol{U}_K^{-1}\sum_{j=1}^K \boldsymbol{D}_j^t\boldsymbol{\Sigma}_j^{-1}\begin{pmatrix} E[(y_j-m_j)(y_i-m_i)] \\ E[\{(y_j-m_j)^2 - \phi_j Q_i(m_j)\}(y_i-m_i)]\end{pmatrix} \\ =& \boldsymbol{e}_0^t\boldsymbol{U}_K^{-1}\boldsymbol{D}_i^t\boldsymbol{\Sigma}_i^{-1}\begin{pmatrix} E[(y_i-m_i)^2] \\ E[\{(y_i-m_j)^2 - \phi_i Q_i(m_i)\}(y_i-m_i)]\end{pmatrix},\end{aligned}$$

which is of order $O(K^{-1})$. Similarly, we can verify that $E[(\hat{\nu}-\nu)(y_i-m_i)^2] = O(K^{-1})$. Therefore, Lemma 4.1 is proved. ■

Thus, from Lemma 4.1, all the results given in Propositions 3.1, 3.2, 3.3 and 3.4 and Theorem 3.1 hold for the estimator $\delta_i^{CEB}(\boldsymbol{y},\widehat{\boldsymbol{\eta}})$. Especialy, we can estimate MSE of CEB using the following parametric bootstrap samples:

$$\begin{aligned} y_i^*|\theta_i^* \sim& f(y_i^*|\theta_i^*) = \exp[n_i(\theta_i^* y_i^* - \psi(\theta_i^*)) + c(y_i^*, n_i)], \\ \theta_i^*|\hat{\nu},\hat{m}_i \sim& \pi(\theta_i^*|\hat{\nu},\hat{m}_i) = \exp[\hat{\nu}(\hat{m}_i\theta_i^* - \psi(\theta_i^*))]C(\hat{\nu},\hat{m}_i),\end{aligned} \tag{4.14}$$

Then, a second-order unbiased estimator of MSE for CEB $\delta_i^{CEB}(\boldsymbol{y},\widehat{\boldsymbol{\eta}})$ is given by

$$mse(\boldsymbol{y},\delta_i^{CEB}(\boldsymbol{y},\widehat{\boldsymbol{\eta}})) = mse^*(\boldsymbol{y},\widehat{\xi}(\boldsymbol{y},\widehat{\boldsymbol{\eta}})) + \widehat{I}_2(\boldsymbol{y}) + 2I_3^*(\boldsymbol{y}), \tag{4.15}$$

where $mse^*(\boldsymbol{y}, \widehat{\xi}(\boldsymbol{y}, \widehat{\boldsymbol{\eta}})) = 2g_1(\widehat{\boldsymbol{\eta}}) - E^*[g_1(\widehat{\boldsymbol{\eta}}^*)] + g_2^*(\widehat{\boldsymbol{\eta}})$, $g_2^*(\widehat{\boldsymbol{\eta}}) = E^*[\{\widehat{\xi}(\boldsymbol{y}^*, \widehat{\boldsymbol{\eta}}^*) - \widehat{\xi}(\boldsymbol{y}^*, \widehat{\boldsymbol{\eta}})\}^2]$, $\widehat{I}_2(\boldsymbol{y}) = \{(a_B(\boldsymbol{y}) - 1)(\widehat{\xi}_i(\boldsymbol{y}, \widehat{\boldsymbol{\eta}}) - \overline{\widehat{\xi}}_w(\boldsymbol{y}, \widehat{\boldsymbol{\eta}})) + \Delta_m(\boldsymbol{y})\}^2$ and

$$I_3^*(\boldsymbol{y}) = E^*[\{\widehat{\xi}_i(\boldsymbol{y}^*, \widehat{\boldsymbol{\eta}}^*) - \widehat{\xi}_i(\boldsymbol{y}^*, \widehat{\boldsymbol{\eta}})\}\{(a_B(\boldsymbol{y}^*) - 1)(\widehat{\xi}_i(\boldsymbol{y}^*, \widehat{\boldsymbol{\eta}}^*) - \overline{\widehat{\xi}}_w(\boldsymbol{y}^*, \widehat{\boldsymbol{\eta}}^*)) + \Delta_m(\boldsymbol{y}^*)\}],$$

for $g_1(\widehat{\boldsymbol{\eta}}) = q_i^2\hat{\nu}[(n_i + \hat{\nu})(\hat{\nu} - v_{2,i})]^{-1}Q_i(\hat{m}_i)$ and $g_1(\widehat{\boldsymbol{\eta}}^*) = q_i^2\hat{\nu}^*[(n_i + \hat{\nu}^*)(\hat{\nu}^* - v_{2,i})]^{-1}Q_i(\hat{m}_i^*)$.

5 Specific mixed models and an Application to Mortality Rates Estimate

5.1 Poisson-gamma mixture model

As an example of mixed models (4.1), we first treat the Poisson-gamma mixture model which are useful for estimation of relative risk in spatial epidemiology. Suppose that a whole region consists of K areas. For $i = 1, \ldots, K$, let y_i be the number of deaths of a specific disease in the i-th area. Let n_i be the expected number of deaths adjusted by age and sex in the i-th area. Assume that $y_1, \ldots, y_K$ are random variables mutually independently distributed as y_i has a Poisson distribution with mean $\lambda_i n_i$, $\mathcal{P}o(\lambda_i n_i)$, where λ_i is an unknown parameter corresponding to disease risk in the i-th area. Then, an unbiased estimator of λ_i is

$$\hat{\lambda}_i^{SMR} = y_i/n_i, \tag{5.1}$$

which is called the Standardized Mortality Ratio (SMR).

Since the variance of $\hat{\lambda}_i^{SMR}$ is $Var(\hat{\lambda}_i^{SMR}) = \lambda_i/n_i$, the variance gets larger for smaller n_i, namely, SMR has a large fluctuation in an area with small n_i. Thus, the following Poisson-gamma model is suggested to fix this undesirable property:

$$\begin{aligned} y_i|\lambda_i \sim& \mathcal{P}o(n_i\lambda_i), \\ \lambda_i \sim& \mathcal{G}a(m_i\nu/n_i, 1/\nu), \end{aligned} \tag{5.2}$$

where $\mathcal{G}a(\alpha, \beta)$ is a Gamma distribution with mean $\alpha\beta$ and variance $\alpha\beta^2$, and m_i and ν are positive hyper-parameters. Expressing these distributions in a natural exponential family, we have

$$\begin{aligned} y_i|\lambda_i \sim& \exp[n_i(y_i n_i^{-1}\log\lambda_i - \lambda_i) + (y_i \log n_i - \log y_i!)], \\ \lambda_i|\nu, m_i \sim& \exp[\{(m_i/n_i)\nu - 1\}\log\lambda_i - \nu\lambda_i + (m_i/n_i)\nu\log\nu]\{\Gamma(m_i\nu/n_i)\}^{-1}\mathrm{d}\lambda_i. \end{aligned} \tag{5.3}$$

Let $\theta_i = n_i^{-1}\log\lambda_i$ with $\mathrm{d}\lambda_i = n_i\exp[n_i\theta_i]\mathrm{d}\theta_i$. Let $\psi(\theta_i) = \exp[n_i\theta_i] = \lambda_i$. Then, the mixed distribution (5.3) is rewritten as

$$\begin{aligned} y_i|\lambda_i \sim& \exp[n_i(y_i\theta_i - \psi(\theta_i)) + (y_i\log n_i - \log y_i!)], \\ \theta_i|\nu, m_i \sim& \exp[\nu(m_i\theta_i - \psi(\theta_i))]n_i^{-1}(\nu/n_i)^{m_i\nu/n_i}\{\Gamma(m_i\nu/n_i)\}^{-1}\mathrm{d}\theta_i, \end{aligned} \tag{5.4}$$

so that this model is in the framework of (4.1) in Section 4 with $v_{0,i} = v_{2,i} = 0$ and $v_{1,i} = n_i$ for $Q_i(m_i)$. It is noted that $\mu_i = \psi'(\theta_i) = n_i\lambda_i$ and $q_i = 1/n_i$ in this case. The Bayes estimators of μ_i and λ_i are

$$\widehat{\mu}_i^B(y_i, m_i, \nu) = \frac{n_i y_i + \nu m_i}{n_i + \nu} \quad \text{and} \quad \hat{\lambda}_i^B(y_i, m_i, \nu) = \frac{y_i + \nu m_i/n_i}{n_i + \nu}. \tag{5.5}$$

We now consider the benchmarking problems for mortality rates. Let $L = \sum_{i=1}^K n_i$ and $\boldsymbol{y} = (y_1, \ldots, y_K)^t$. Consider the case of $w_i = n_i / \sum_{j=1}^K n_j$. Let $\overline{y}_w = L^{-1} \sum_{i=1}^K y_i$, which is the average number of patients with the disease in the whole reagion. Note that $E[\overline{y}_w] = L^{-1} \sum_{i=1}^K E[y_i] = L^{-1} \sum_{i=1}^K \lambda_i n_i$. This suggests that it is reasonable to consider estimators $\hat{\lambda}_i^C$'s satisfying the mean constraint

(MC) $\sum_{i=1}^K w_i \hat{\lambda}_i^C = L^{-1} \sum_{i=1}^K y_i$, for $L = \sum_{i=1}^K n_i$,

which is the constraint (C1′). It is clear that the SMR $\hat{\lambda}_i^{SMR} = y_i / n_i$ satisfies the constraint. However, the Bayes estimator $\hat{\lambda}_i^B(y_i, m_i, \nu)$ does not satisfy the constraint (MC), namely,

$$L^{-1} \sum_{i=1}^K \frac{n_i y_i + \nu m_i}{n_i + \nu} \neq L^{-1} \sum_{i=1}^K y_i.$$

Then from (4.11) and (4.6), CEB of λ_i under (MC) is

$$\hat{\lambda}_i^{CEBm} = \hat{\lambda}_i^{EB} + \sum_{j=1}^K \frac{w_j}{n_j} \frac{\hat{\nu}(y_j - \hat{m}_j)}{n_j + \hat{\nu}}, \tag{5.6}$$

where $\hat{\lambda}_i^{EB}$ is EB $\hat{\lambda}_i^{EB} = (y_i + \hat{\nu} \hat{m}_i / n_i)/(n_i + \hat{\nu})$.

The variance constraint (C2′) is described as

(VC) $\displaystyle\sum_{i=1}^K w_i (\hat{\lambda}_i^C - \overline{\hat{\lambda}}_w^C)^2 = \sum_{i=1}^K w_i (\hat{\lambda}_i^{EB} - \overline{\hat{\lambda}}_w^{EB})^2 + \frac{1}{K^r} \sum_{i=1}^K w_i (1 - w_i) \frac{\hat{\lambda}_i^{EB}}{n_i + \hat{\nu}}$,

where $\overline{\hat{\lambda}}_w^C = \sum_{j=1}^K w_j \hat{\lambda}_j^C$ and $\overline{\hat{\lambda}}_w^{EB} = \sum_{j=1}^K w_j \hat{\lambda}_j^{EB}$.

It is noted that Frey and Cressie (2003) treated a similar variance constraint with the same weights w_i, where their prior distribution of λ_i is different from ours. Under (VC), from (4.11), CEB is

$$\hat{\lambda}_i^{CEBv} = \hat{\lambda}_i^{EB} + \{a_B(\boldsymbol{y}) - 1\} \{\hat{\lambda}_i^{EB} - \overline{\hat{\lambda}}_w^{EB}\}, \tag{5.7}$$

where

$$\{a_B(\boldsymbol{y})\}^2 = 1 + \frac{1}{K^r} \frac{\sum_{j=1}^K w_j (1 - w_j) \hat{\lambda}_j^{EB} / (n_j + \hat{\nu})}{\sum_{j=1}^K w_j (\hat{\lambda}_j^{EB} - \overline{\hat{\lambda}}_w^{EB})^2}.$$

When both constraints (MC) and (VC) are imposed, it follows from (4.11) that CEB under the mean-variance constraints is

$$\hat{\lambda}_i^{CEBmv} = \hat{\lambda}_i^{EB} + \{a_B(\boldsymbol{y}) - 1\} \{\hat{\lambda}_i^{EB} - \overline{\hat{\lambda}}_w^{EB}\} + \sum_{j=1}^K \frac{w_j}{n_j} \frac{\nu(y_j - \hat{m}_j)}{n_j + \hat{\nu}}. \tag{5.8}$$

Since $m_i = \psi'(\boldsymbol{x}_i^t \boldsymbol{\beta})$ and $\psi(\theta_i) = \exp[n_i \theta_i] = \lambda_i$, it is seen that $m_i = n_i \exp[n_i \boldsymbol{x}_i \boldsymbol{\beta}]$. The unknown parameters are $\boldsymbol{\eta}^t = (\boldsymbol{\beta}^t, \nu)$, and are estimated by the estimating equations (4.6). In this model, $Q_i(m_i) = n_i m_i$, $\phi_i = 1/\nu + 1/n_i$, $v_{2,i} = 0$, $g_{1i} = y_i - m_i$ and

$g_{2i} = (y_i - m_i)^2 - m_i(1+\tau_i)$ for $\tau_i = n_i/\nu$. For $\mu_{ri} = E[(y_i - m_i)^r]$, $i = 2, 3, 4$, from Proposition 4.1, it is observed that

$$\begin{aligned}\mu_{2i} =& m_i(1+\tau_i), \quad \mu_{3i} = m_i(1+3\tau_i+2\tau_i^2),\\ \mu_{4i} =& m_i\{1+3m_i+(6m_i+7)\tau_i+3(m_i+4)\tau_i^2+6\tau_i^3\}.\end{aligned}$$

Hence, $(\boldsymbol{\beta}, \nu)$ can be estimated by solving the estimating equations

$$\begin{aligned}&\sum_{i=1}^K \frac{1}{|\boldsymbol{\Sigma}_i|}\Big[\{\mu_{4i} - \mu_{2i}^2 - \mu_{3i}(1+\tau_i)\}g_{1i} + \{\mu_{2i}(1+\tau_i) - \mu_{3i}\}g_{2i}\Big]n_i m_i \boldsymbol{x}_i = \boldsymbol{0},\\ &\sum_{i=1}^K \frac{1}{|\boldsymbol{\Sigma}_i|}\{\mu_{2i}g_{2i} - \mu_{3i}g_{1i}\}n_i m_i = 0.\end{aligned}$$

The asymptotically unbiased estimators of MSE of CEB $\hat{\lambda}_i^{CEBm}$, $\hat{\lambda}_i^{CEBv}$ and $\hat{\lambda}_i^{CEBmv}$ are provided by (4.15), where $g_1(\boldsymbol{\eta})$ in $mse^*(\boldsymbol{y}, \hat{\lambda}_i)$ is given by $g_1(\boldsymbol{\eta}) = m_i/\{n_i(n_i+\nu)\}$.

5.2 An example for benchmarking SMR in mortality rates estimates

We now apply the procedures given in Section 5.1 to real mortality data, and we investigate how CEB works.

The dataset consists of observed number of mortality y_i and its expected number n_i of stomach cancer for females lived in the i-th city or town in Saitama prefecture, Japan, for five years from 1995 to 1999. Saitama is the prefecture next to Tokyo. Such area level data are available for $K = 92$ cities and towns, and the total number of mortality in the whole region is $L = 3,953$. The expected numbers are adjusted by age based on the population so that $L = \sum_{i=1}^K y_i = \sum_{i=1}^K n_i$. We apply the Poisson-gamma model (5.2) to this dataset and estimate the stomach cancer mortality risk λ_i for each municipality. For the regression part $\boldsymbol{x}_i^t\boldsymbol{\beta}$, we here consider the case that $\boldsymbol{x}_i$ is one-dimensional and $\boldsymbol{x}_i = 1$, namely, $m_i = n_i \exp[n_i\beta_0]$ for $\boldsymbol{\beta} = \beta_0$.

Using the package SolveNLE in the programming language Ox for solving non-linear equations, we can solve the estimating equations (4.6) and get the estimates $\beta_0 = 1.53249 \times 10^{-4}$ and $\nu = 174.472$. It is observed that $L = \sum_{i=1}^K n_i = 3,953$, and that for $w_i = n_i/L$, $\sum_{i=1}^K w_i\hat{\lambda}_i^{SMR} = 1$ and $\sum_{i=1}^K w_i\hat{\lambda}_i^{EB} = 1.0086$, which means that their difference is quite small. On the other hand, the weighted sample variance of SMR is $\sum_{i=1}^K w_i(\hat{\lambda}_i^{SMR} - 1)^2 = 0.0302$, while the weighted sample variance of the empirical Bayes estimates (EB) is 0.0014. However, the right hand side in (VC) with $r = 0$ in Section 5.1 is

$$\sum_{i=1}^K w_i(\hat{\lambda}_i^{EB} - \overline{\hat{\lambda}}_w^{EB})^2 + \sum_{i=1}^K w_i(1-w_i)\frac{\hat{\lambda}_i^{EB}}{n_i + \hat{\nu}} = 0.0051,$$

this implies that EB yields over-shrinking. Thus, we investigate the behavior of CEB under the mean-variance constraints with $r = 0$ and $a_B(\boldsymbol{y}) = 1.9242$, and we can confirm

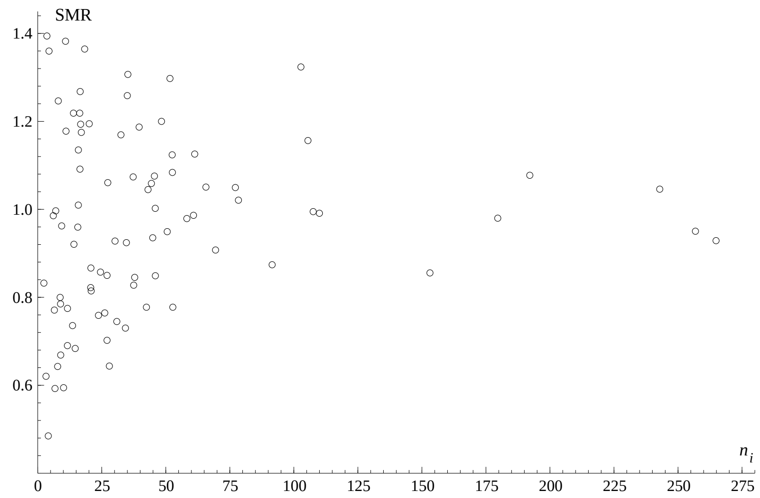

Figure 1: Plots of SMR for stomach cancer mortality incidence of females in Saitama prefecture

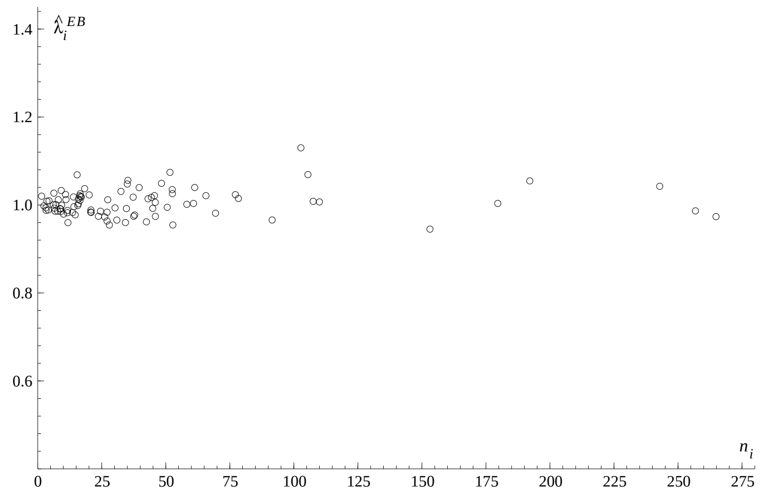

Figure 2: Plots of EB for stomach cancer mortality incidence of females in Saitama prefecture

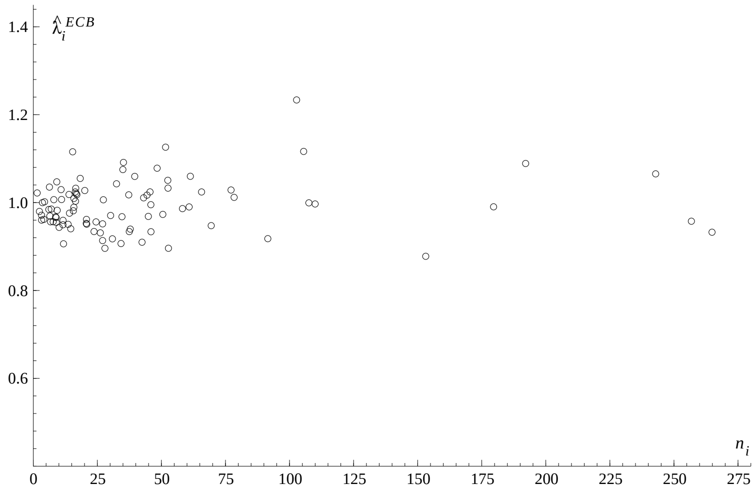

Figure 3: Plots of CEB under mean-variance constraints with $r = 0$ for stomach cancer mortality incidence of females in Saitama prefecture

that the weighted sample variance of CEB is identical to 0.0051. Plots of estimates by SMR, EB and CEB are illustrated in Figures 1, 2 and 3,respectively. These figures show that EB shrinks SMR too much and that CEB expands slightly shrunken EB.

For 92 municipalities in Saitama prefecture, Table 1 reports the values of SMR $\hat{\lambda}_i^{SMR}$, EB $\hat{\lambda}_i^{EB}$ and two kinds of CEB under mean-variance constraints with $r = 0$ and $r = 0.5$, denoted by CEB_0 and $\text{CEB}_{0.5}$, and estimates of their MSEs by the parametric bootstrap method, where the estimates are constructed based on 1,000 bootstrap samples. It is noted that the columns '*estimates*' and '*estimated MSE*' report the values of $100 \times \hat{\lambda}_i$ and $100 \times mse(\hat{\lambda}_i)$, where $mse(\hat{\lambda}_i)$ is an estimate of MSE of $\hat{\lambda}_i$. It is noted that an estimator of MSE of SMR, y_i/n_i, is provided by the parametric bootstrap method. In fact, note that $E[(y_i/n_i - \lambda_i)^2] = n_i^{-2}E[(y_i - n_i\lambda_i)^2] = n_i^{-1}E[\lambda_i] = m_i/n_i^2 = n_i^{-1}\exp\{n_i\beta_0\}$. Then, the second order unbiased estimator of MSE of SMR is given as

$$mse(\hat{\lambda}_i^{SMR}) = 2n_i^{-1}\exp\{n_i\widehat{\beta}_0\} - n_i^{-1}E^*[\exp\{n_i\widehat{\beta}_0^*\}].$$

For smaller n_i, SMR has larger *mse*, while EB, CEB_0 and $\text{CEB}_{0.5}$ have smaller *mse*. EB shrinks SMR very much, and CEB_0 and $\text{CEB}_{0.5}$ expand shrunken EB slightly. For larger n_i, on the other hand, EB, CEB_0 and $\text{CEB}_{0.5}$ do not shrink SMR so much, but expand shrunken EB slightly. For most cities and towns, SMR has larger *mse*, while EB, CEB_0 and $\text{CEB}_{0.5}$ have smaller *mse*. For Kawagoe, Kawaguchi, Urawa, Oomiya and Tokorozawa, which have n_i's larger than 179, their *mse* values of EB, CEB_0 and $\text{CEB}_{0.5}$ are slightly larger than those of SMR. For Kumagaya with $n_i = 102.7$, *mse* of CEB_0 is not good, while *mse* of $\text{EB}_{0.5}$ is close to *mse* of EB. Also for Kasukabe with $n_i = 105.5$, *mse* of CEB_0 is close to *mse* of SMR. For $\text{CEB}_{0.5}$, the values of the estimates and the estimated MSE are very close to the values for SMR. Taking into account the motivation of Louis (1984) and Ghosh (1992), we suggest the use of CEB_0, but we need to care the values of estimates in the cases of large *mse* (estimated MSE). For example, the estimate 1.23 for Kumagaya should be noted because it has the large *mse*.

5.3 Binomial-beta mixture and logistic regression

Finally, we give a breif explanation of the beta-binomial model with logistic regression. For $i = 1, \ldots, K$, let y_i be the number of counts of a specific event in n_i trials for the i-th point. Assume that $y_1, \ldots, y_K$ are random variables mutually independently distributed as $n_i y_i$ has a binomial distribution $\mathcal{B}in(n_i, p_i)$, and p_i is a probability of occurrence of the event having the beta distribution. Then, the binomial-beta mixture is given by

$$\begin{aligned} y_i|p_i \sim& \mathcal{B}in(n_i, p_i), \\ p_i \sim& beta(m_i n_i^{-1}\nu, (1 - m_i n_i^{-1})\nu), \end{aligned} \tag{5.9}$$

where m_i and ν are positive hyper-parameters. Expressing these distributions in a natural exponential family, we have

$$\begin{aligned} y_i|p_i \sim& \exp\Big[n_i\big\{y_i n_i^{-1}\log\big(\frac{p_i}{1-p_i}\big) + \log(1-p_i)\big\}\Big]\begin{pmatrix} n_i \\ y_i \end{pmatrix}, \\ p_i|\nu, m_i \sim& \exp\Big[m_i n_i^{-1}\nu\log\big(\frac{p_i}{1-p_i}\big) + \nu\log(1-p_i)\Big]\frac{p_i^{-1}(1-p_i)^{-1}\Gamma(\nu)}{\Gamma(m_i n_i^{-1}\nu)\Gamma((1-m_i n_i^{-1})\nu)}\mathrm{d}p_i. \end{aligned} \tag{5.10}$$

Let $\theta_i = n_i^{-1}\log(p_i/(1-p_i))$ with $p_i^{-1}(1-p_i)^{-1}\mathrm{d}p_i = n_i \mathrm{d}\theta_i$. Let $\psi(\theta_i) = \log(1+\exp[n_i\theta_i]) = -\log(1-p_i)$. Then, the mixed distribution (5.10) is rewritten as

$$\begin{aligned} y_i|p_i \sim& \exp\Big[n_i(y_i\theta_i - \psi(\theta_i))\Big]\begin{pmatrix} n_i \\ y_i \end{pmatrix}, \\ \theta_i|\nu, m_i \sim& \exp\Big[\nu(m_i\theta_i - \psi(\theta_i))\Big]\frac{\Gamma(\nu)n_i}{\Gamma(m_i n_i^{-1}\nu)\Gamma((1-m_i n_i^{-1})\nu)}\mathrm{d}\theta_i, \end{aligned} \tag{5.11}$$

so that this model is in the framework of (4.1) in Section 4 with $v_{0,i} = 0$, $v_{1,i} = n_i$ and $v_{2,i} = -1$ for $Q_i(\cdot)$. Since $\mu_i = \psi'(\theta_i) = n_i p_i$, the Bayes estimators of μ_i and p_i are

$$\widehat{\mu}_i(y_i, m_i, \nu) = \frac{n_i y_i + \nu m_i}{n_i + \nu} \quad \text{and} \quad \hat{p}_i(y_i, m_i, \nu) = \frac{y_i + \nu m_i/n_i}{n_i + \nu}.$$

Since $m_i = \psi'(\boldsymbol{x}_i^t\boldsymbol{\beta})$ and $\psi(\theta_i) = \log(1 + \exp[n_i\theta_i]) = -\log(1-p_i)$, it is seen that $m_i = n_i\exp[n_i\boldsymbol{x}_i^t\boldsymbol{\beta}]/(1 + \exp[n_i\boldsymbol{x}_i^t\boldsymbol{\beta}])$. The unknown parameters are $\boldsymbol{\eta}^t = (\boldsymbol{\beta}^t, \nu)$, and are estimated by the estimating equations (4.6). In this model, $Q_i(m_i) = n_i m_i - m_i^2$, $\phi_i = (1+\nu/n_i)/(\nu+1)$, $v_{2,i} = -1$, $g_{1i} = y_i - m_i$ and $g_{2i} = (y_i - m_i)^2 - m_i(n_i - m_i)\phi_i$.

For $\mu_{ri} = E[(y_i - m_i)^r]$, $i = 2, 3, 4$, from Proposition 4.1, it is observed that

$$
\begin{aligned}
\mu_{2i} =& \frac{\nu/n_i + 1}{\nu + 1} m_i(n_i - m_i), \quad \mu_{3i} = \frac{(\nu/n_i + 1)(\nu/n_i + 2)}{(\nu + 1)(\nu + 2)} m_i(n_i - m_i)(n_i - 2m_i),\\
\mu_{4i} =& m_i^2(n_i - m_i)^2 \Big[3\frac{(d_i + 1)(2d_i + 1)(3d_i + 1)}{(\nu + 1)(\nu + 2)(\nu + 3)} \big\{ \nu + 2 + 2\frac{(n_i - 2m_i)^2}{m_i(n_i - m_i)} \big\}\\
&+ \frac{12}{n_i} \frac{(d_i + 1)(2d_i + 1)}{(\nu + 1)(\nu + 2)} \frac{(n_i - 2m_i)^2}{m_i(n_i - m_i)} + \frac{1}{n_i^2} \frac{d_i + 1}{\nu + 1} \big\{ 7\frac{(n_i - 2m_i)^2}{m_i(n_i - m_i)} + 2n_i(4d_i + 3) \big\}\\
&+ \frac{2d_i + 3}{n_i^2} + \frac{1}{n_i^3} \frac{(n_i - 2m_i)^2}{m_i(n_i - m_i)} \Big].
\end{aligned}
$$

Thus, $(\boldsymbol{\beta}, \nu)$ can be estimated by the estimating equations (4.6).

We can treat the benchmarking problems which have been studied in Section 5.1. Although the details are omitted here, CEB is provided for the mean constraint (MC) and/or the variance constraint (VC), and second-order unbiased estimators of their MSE are provided.

6 Concluding Remarks

In this paper, we have considered the constrained empirical Bayes (CEB) estimators under the mean and/or variance constriants and derived asymptotic approximations of MSE of CEB in the general mixed models. As pointed out by Louis (1984), the sample variance in the empirical Bayes (EB) estimates is smaller than the posterior variance, and Ghosh (1992) suggested CEB so that the sample variance in CEB is identical to the posterior variance. Then, it has been shown that MSE of CEB is larger than MSE of EB in the first order approximation. Thus, it is important to assess uncertainty of CEB, namely, estimation of MSE of CEB. We have provided a second-order unbiased estimator for MSE of CEB.

These general results have been applied to natural exponential families with quadratic variance functions (NEF-QVF) based on the study made by Morris (1982, 83) and Ghosh and Maiti (2004). The application includes Poisson-gamma and binomial-beta mixture models, and we have analized real mortality data of stomach cancer for female in cities or towns in Saitama prefecture. Through this example, it is found that EB yields over-shrinking, while CEB under the mean-variance constraints expands EB slightly. It is also found that the estimated MSEs of CEB are larger than those of SMR for cities with larger expected numbers of mortality, while the estimated MSE of CEB are much smaller for most cities and towns. This is an interesting phenomenon and tells us the importance of estimate of MSE for CEB.

Acknowledgments.

The research of the first author was supported in part by Grant-in-Aid for Scientific Research Nos. 21540114 and 23243039 from Japan Society for the Promotion of Science.

References

[1] Breslow, N.E., and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *J. American Statist. Assoc.*, **88**, 9-25.

[2] Breslow, N.E., and Lin, X. (1995). Bias correction in generalized linear mixed models with a single component of dispersion. *Biometrika*, **82**, 81-91.

[3] Butar, F.B. and Lahiri, P. (2003). On measures of uncertainty of empirical Bayes small-area estimators. *J. Statist. Plan. Inf.*, **112**, 63-76.

[4] Clayton, D. and Kaldor, J.(1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, **43**, 671-681.

[5] Crescenzi, M., Ghosh, M. and Maiti, T. (2005). Empirical Bayes estimators with uncertainty measures for NEF-QVF populations. *JIRSS*, **4**, 1-19.

[6] Datta, G.S., Ghosh, M., Steorts, R., and Maples, J. (2011). Bayesian benchmarking with applications to small area estimation. *Test*, **20**, 574-588.

[7] Frey, J., and Cressie, N. (2003). Some results on constrained Bayes estimators. *Statist. Probab. Letters*, **65**, 389-399.

[8] Ghosh, M. (1992). Constrained Bayes estimation with applications. *J. American Statist. Assoc.*, **87**, 533-540.

[9] 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.

[10] 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 J. Statist.*, **35**, 484-495.

[11] Godambe, V.P. and Thompson, M.E. (1989). An extension of quasi-likelihood estimation (with Discussion). *J. Statist. Plan. Infer.*, **22**, 137-152.

[12] Kubokawa, T. (2011). On measuring uncertainty of small area estimators with higher order accuracy. *J. Japan Statist. Soc.*, **41**, 93-119.

[13] Kubokawa, T. (2012) Mixed effects prediction under benchmarking and applications to small area estimation. Discussion Paper Series, CIRJE-F-832.

[14] Louis, T.A. (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods. *J. Amer. Statist. Assoc.*, **79**, 393-398.

[15] Manton, K.G., Woodbury, M.A., Stallard, E., Riggan, W.B., Creason, J.P. and Pellom, A.C.(1989). Empirical Bayes procedures for stabilizing maps of U.S. cancer mortality rates. *J. American Statist. Assoc.*, **84**, 637-649.

[16] Morris, C. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.*, **10**, 65-80.

[17] Morris, C. (1983). Natural exponential families with quadratic variance functions: statistical theory. *Ann. Statist.*, **11**, 515-529.

[18] Sarkar, S. and Ghosh, M. (1998). Empirical Bayes estimation of local area means for NEF-QVF superpopulations. *Sankhya*, **B 60**, 464-487.

[19] Tango, T. (1988). Empirical Bayes estimation for mortality indices ; application to disease mapping. *Japanese J. Applied Statist.*, **17**, 81-96 (in Japanese).

[20] Tsutakawa, R.K. (1985). Estimation of cancer mortality rates: A Bayesian analysis of small frequencies. *Biometrics*, **41**, 69-79.

[21] Tsutakawa, R.K. (1988). Mixed model for analyzing geographic variability in mortality rates. *J. American Statist. Assoc.*, **83**, 37-42.

[22] Tsutakawa, R.K., Shoop, G.L. and Marienfeld, C.J. (1985). Empirical Bayes estimation of cancer mortality rates. *Statist. Medicine*, **4**, 201-212.

A Appendix

We here provide the proof of Proposition 4.1. Throughout the appendix, we omit the index i in y_i, μ_i, θ_i and others. Note that $E[y-\mu]=0$ for $\mu=\psi'(\theta)$. It follows that $E[(y-\mu)^2|\theta]=n^{-1}\psi''(\theta)$, $E[(y-\mu)^3|\theta]=n^{-2}\psi'''(\theta)$ and $E[(y-\mu)^4|\theta]=3n^{-2}\{\psi''(\theta)\}^2+n^{-3}\psi^{(4)}(\theta)$. Also, note that $\psi'(\theta)=\mu$, $\psi''(\theta)=Q(\mu)=v_0+v_1\mu+v_2\mu^2$, $\psi'''(\theta)=Q(\mu)Q'(\mu)$ and $\psi^{(4)}(\theta)=Q(\mu)\{Q'(\mu)\}^2+2v_2\{Q(\mu)\}^2$. Thus, we have

$$\begin{aligned}
E[(y-\mu)^2|\theta] =& n^{-1}Q(\mu),\\
E[(y-\mu)^3|\theta] =& n^{-2}Q(\mu)Q'(\mu),\\
E[(y-\mu)^4|\theta] =& (3n^{-2}+2v_2n^{-3})\{Q(\mu)\}^2+n^{-3}Q(\mu)\{Q'(\mu)\}^2.
\end{aligned} \tag{A.1}$$

Letting $T=\mu-m$, we can express $Q(\mu)$ and $Q'(\mu)$ as $Q(\mu)=Q(m)+Q'(m)T+v_2T^2$ and $Q'(\mu)=Q'(m)+2v_2T$, which are used to rewrite (A.1) as

$$\begin{aligned}
E[(y-\mu)^2|\theta] =& n^{-1}\{Q(m)+Q'(m)T+v_2T^2\},\\
E[(y-\mu)^3|\theta] =& n^{-2}\{Q(m)+Q'(m)T+v_2T^2\}\{Q'(m)+2v_2T\},\\
E[(y-\mu)^4|\theta] =& (3n^{-2}+2v_2n^{-3})\{Q(m)+Q'(m)T+v_2T^2\}^2\\
&+n^{-3}\{Q(m)+Q'(m)T+v_2T^2\}\{Q'(m)+2v_2T\}^2.
\end{aligned} \tag{A.2}$$

By integration by parts, we have the equations $E[T^2]=Q(m)/\nu+(v_2/\nu)E[T^2]$, $E[T^3]=(2/\nu)Q'(m)E[T^2]+(2/\nu)v_2E[T^3]$ and $E[T^4]=(3/\nu)Q(m)E[T^2]+(3/\nu)Q'(m)E[T^3]+$

$(3/\nu)v_2E[T^4]$, which imply that

$$
\begin{aligned}
E[T^2] =& \frac{Q(m)}{\nu - v_2}, \\
E[T^3] =& \frac{2Q(m)Q'(m)}{(\nu - v_2)(\nu - 2v_2)}, \\
E[T^4] =& \frac{3Q(m)\{(\nu - 2v_2)Q(m) + 2\{Q'(m)\}^2\}}{(\nu - v_2)(\nu - 2v_2)(\nu - 3v_2)}.
\end{aligned}
\tag{A.3}
$$

Based on these observations, we can derive the moments given in Proposition 4.1. Since $E[\mu - m] = 0$, it is clear that $E[y - m] = 0$. Since $E[(y - m)^2] = E[\{(y - \mu) + (\mu - m)\}^2] = E[(y - \mu)^2] + E[(\mu - m)^2]$, it is seen that $E[(y - m)^2] = n^{-1}E[Q(m) + Q'(m)T + v_2T^2] + Q(m)/(\nu - v_2)$, which yields that $E[(y - m)^2] = Q(m)(1 + \nu/n)/(\nu - v_2)$. Since $E[(y - m)^3] = E[(y - \mu)^3 + 3(y - \mu)^2(\mu - m) + (\mu - m)^3]$, it is seen that

$$
\begin{aligned}
E[(y - m)^3] =& n^{-2}E[\{Q(m) + Q'(m)T + v_2T^2\}\{Q'(m) + 2v_2T\}] \\
&+ 3n^{-1}E[\{Q(m) + Q'(m)T + v_2T^2\}T] + E[T^3] \\
=& (1 + 3d + 2d^2)E[T^3] + \frac{3}{n}(1 + d)Q'(m)E[T^2] + n^{-2}Q(m)Q'(m) \\
=& \frac{Q(m)Q'(m)}{(\nu - v_2)(\nu - 2v_2)}\big\{2 + 6d + 4d^2 + 3(1 + d)(\frac{\nu}{n} - 2d) + (\frac{\nu}{n} - d)(\frac{\nu}{n} - 2d)\big\},
\end{aligned}
$$

for $d = v_2/n$. This is summarized as

$$
E[(y - m)^3] = \frac{Q(m)Q'(m)}{(\nu - v_2)(\nu - 2v_2)}(\frac{\nu}{n} + 1)(\frac{\nu}{n} + 2).
$$

Similarly, $E[(y - m)^4] = E[(y - \mu)^4 + 4(y - \mu)^3(\mu - m) + 6(y - \mu)^2(\mu - m)^2 + (\mu - m)^4]$, so that

$$
\begin{aligned}
E[(y - m)^4] =& (3n^{-2} + 2v_2n^{-3})E[\{Q(m) + Q'(m)T + v_2T^2\}^2] \\
&+ n^{-3}E[\{Q(m) + Q'(m)T + v_2T^2\}\{Q'(m) + 2v_2T\}^2] \\
&+ 4n^{-2}E[\{Q(m) + Q'(m)T + v_2T^2\}\{Q'(m) + 2v_2T\}T] \\
&+ 6n^{-1}E[\{Q(m) + Q'(m)T + v_2T^2\}T^2] + E[T^4] \\
=& \{6d^3 + 1d^2 + 6d + 1\}E[T^4] + \frac{6}{n}\{2d^2 + 3d + 1\}Q'(m)E[T^3] \\
&+ \Big\{\frac{7}{n^2}(d + 1)\{Q'(m)\}^2 + \frac{2}{n}(4d^2 + 7d + 3)Q(m)\Big\}E[T^2] \\
&+ \frac{1}{n^3}Q(m)\{n(3 + 2d)Q(m) + \{Q'(m)\}^2\},
\end{aligned}
$$

which is summarized as

$$
\begin{aligned}
E[(y - m)^4] =& (d + 1)(2d + 1)(3d + 1)E[T^4] + \frac{6}{n}(d + 1)(2d + 1)Q'(m)E[T^3] \\
&+ \frac{1}{n^2}(d + 1)\Big\{7\{Q'(m)\}^2 + 2n(4d + 3)Q(m)\Big\}E[T^2] \\
&+ \frac{1}{n^3}Q(m)\{n(3 + 2d)Q(m) + \{Q'(m)\}^2\}.
\end{aligned}
$$

Hence, we get the expression given in Proposition 4.1.

Table 1: Estimates by SMR, EB and CEB under mean-variance constraints, and estimates of their MSE for 92 cities and towns in Saitama prefecture (n_i is the expected number of female mortality by stomach cancer, CEB_0 and $CEB_{0.5}$ correspond the cases of $r = 0$ and $r = 0.5$, respectively. $100 \times \hat{\lambda}_i$ and $100 \times mse(\hat{\lambda}_i)$ are given as *estimates* and *estimated MSE*.)

		estimates				estimated MSE			
city, town	n_i	SMR	EB	CEB$_0$	CEB$_{0.5}$	SMR	EB	CEB$_0$	CEB$_{0.5}$
Kawagoe	192.1	107	105	108	105	0.536	0.561	0.678	0.562
Kumagaya	102.7	132	112	123	113	0.989	0.732	1.804	0.738
Kawaguchi	242.8	104	104	106	103	0.427	0.497	0.549	0.499
Urawa	256.7	95	98	95	97	0.405	0.482	0.565	0.495
Oomiya	264.8	92	97	93	96	0.393	0.474	0.641	0.491
Gyouda	61.2	112	103	105	103	1.647	0.855	0.896	0.857
Chichibu	50.5	94	99	97	98	1.993	0.895	0.941	0.906
Tokorozawa	179.6	97	100	99	99	0.572	0.580	0.597	0.589
Hannou	58.2	97	100	98	99	1.733	0.866	0.889	0.875
Kazo	44.3	105	101	101	100	2.268	0.919	0.919	0.925
Honjyo	45.9	84	97	93	96	2.192	0.913	1.076	0.930
Higashimatsuyama	52.7	77	95	89	93	1.911	0.886	1.228	0.911
Iwatsuki	65.6	105	102	102	101	1.538	0.840	0.841	0.845
Kasukabe	105.5	115	106	111	106	0.963	0.725	0.949	0.725
Sayama	91.5	87	96	91	95	1.108	0.762	0.992	0.782
Hanyu	42.4	77	96	90	94	2.371	0.927	1.196	0.949
Kounosu	45.5	107	102	102	101	2.210	0.914	0.915	0.919
Fukaya	69.4	90	98	94	96	1.456	0.828	0.942	0.843
Ageo	109.9	99	100	99	99	0.925	0.714	0.725	0.722
Yono	48.3	120	104	107	104	2.084	0.903	0.987	0.904
Souka	107.5	99	100	99	99	0.945	0.720	0.728	0.728
Koshigaya	153.1	85	94	87	92	0.668	0.624	1.077	0.653
Warabi	45.9	100	100	99	99	2.194	0.913	0.925	0.921
Toda	44.9	93	99	96	98	2.242	0.917	0.973	0.928
Iruma	77.1	104	102	102	101	1.311	0.803	0.806	0.808
Hatogaya	35.2	130	105	109	105	2.856	0.958	1.082	0.958
Asaka	52.4	112	103	105	102	1.920	0.887	0.912	0.890
Shiki	32.4	116	103	104	102	3.093	0.970	0.984	0.973
Wakou	30.1	92	99	97	98	3.329	0.981	1.032	0.992
Niiza	78.3	102	101	101	100	1.291	0.800	0.801	0.806
Okegawa	43.0	104	101	101	100	2.337	0.924	0.926	0.931

		estimates				estimated MSE			
city, town	n_i	SMR	EB	CEB_0	$CEB_{0.5}$	SMR	EB	CEB_0	$CEB_{0.5}$
Kuki	39.5	118	103	105	103	2.541	0.939	0.979	0.941
Kitamoto	37.2	107	101	101	101	2.700	0.949	0.949	0.954
Yashio	37.4	82	97	93	96	2.685	0.948	1.110	0.965
Fujimi	52.5	108	102	103	101	1.917	0.887	0.892	0.891
Kamifukuoka	34.2	73	96	90	94	2.935	0.962	1.247	0.985
Misato-shi	60.8	98	100	99	99	1.659	0.857	0.875	0.866
Hasuda	37.8	84	97	93	96	2.657	0.946	1.087	0.962
Sakado	51.6	129	107	112	107	1.952	0.891	1.161	0.891
Satte	34.9	125	104	107	104	2.876	0.959	1.035	0.960
Tsurugashima	30.8	74	96	91	95	3.254	0.977	1.210	0.998
Hidaka	34.6	92	99	96	98	2.903	0.961	1.018	0.972
Yoshikawa	27.9	64	95	89	93	3.591	0.991	1.334	1.016
Ina	15.6	95	99	98	98	6.412	1.053	1.084	1.063
Fukiage	17.0	117	101	101	101	5.890	1.046	1.046	1.051
Ooi	20.7	86	98	96	97	4.830	1.026	1.099	1.039
Miyoshi	20.6	82	98	95	97	4.851	1.027	1.125	1.041
Moroyama	23.7	75	97	93	96	4.231	1.012	1.173	1.029
Ogose	10.8	138	102	102	101	9.229	1.080	1.083	1.084
Naguri	3.5	139	100	99	99	27.898	1.122	1.130	1.130
Namekawa	8.9	66	98	95	97	11.161	1.090	1.182	1.104
Arashiyama	14.1	92	99	97	98	7.096	1.061	1.102	1.072
Ogawa	27.0	70	96	91	94	3.711	0.995	1.246	1.016
Tokigawa	6.7	59	98	95	97	14.830	1.103	1.191	1.117
Tamakawa	4.1	48	98	96	97	24.257	1.119	1.191	1.132
Kawashima	16.5	126	102	103	101	6.053	1.048	1.053	1.052
Yoshimi	15.3	182	106	111	106	6.517	1.055	1.275	1.055
Hatoyama	13.5	73	98	95	97	7.371	1.064	1.169	1.079
Yokose	7.7	64	98	95	97	12.870	1.097	1.184	1.111
Minano	11.5	69	98	94	97	8.642	1.075	1.183	1.090
Nagatoro	8.0	124	101	100	100	12.480	1.096	1.099	1.102
Yoshida	6.4	77	99	96	98	15.434	1.105	1.159	1.116

		estimates				estimated MSE			
city, town	n_i	SMR	EB	CEB_0	$\mathrm{CEB}_{0.5}$	SMR	EB	CEB_0	$\mathrm{CEB}_{0.5}$
Okano	11.7	33	95	90	94	8.491	1.074	1.361	1.097
ryoujin	3.2	30	98	96	97	30.852	1.125	1.201	1.137
Ootaki	2.4	83	99	97	98	41.626	1.130	1.163	1.140
Arakawa	6.0	98	100	98	99	16.440	1.107	1.133	1.117
Higashichichibu	4.4	135	100	100	100	22.678	1.117	1.123	1.124
Misato-machi	11.0	117	101	100	100	9.075	1.079	1.081	1.085
Kodama	16.4	121	102	102	101	6.108	1.049	1.050	1.054
Kamikawa	10.0	59	97	94	96	9.923	1.084	1.211	1.099
Kamiizumi	1.5	328	101	102	101	65.719	1.135	1.136	1.140
Kamisato	18.3	136	103	105	103	5.473	1.039	1.071	1.041
Oosoto	6.3	174	102	103	102	15.864	1.106	1.113	1.110
Kounan	9.3	96	99	98	98	10.707	1.088	1.117	1.098
Menuma	20.8	81	98	95	97	4.807	1.026	1.129	1.040
Okabe	13.9	121	101	101	101	7.184	1.062	1.062	1.068
Kawamoto	9.1	163	103	104	102	10.913	1.089	1.109	1.092
Hanazono	8.7	79	99	96	98	11.440	1.092	1.150	1.103
Yorii	27.3	106	101	100	100	3.673	0.994	0.997	1.001
Kisai	14.6	68	97	94	96	6.851	1.059	1.197	1.075
Minamikawara	3.2	62	99	97	98	31.038	1.125	1.175	1.136
Kawazato	7.0	99	100	98	99	14.253	1.102	1.126	1.111
Kitakawabe	8.9	78	99	96	97	11.227	1.091	1.153	1.103
Ootone	11.6	77	98	95	97	8.622	1.075	1.153	1.088
Miyashiro	20.0	119	102	102	101	4.992	1.030	1.032	1.034
Shiraoka	26.1	76	97	93	95	3.837	1.000	1.173	1.017
Shoubu	15.8	113	101	100	100	6.321	1.052	1.054	1.058
Kurihashi	15.8	100	100	98	99	6.325	1.052	1.071	1.061
Washimiya	16.7	119	101	102	101	5.983	1.047	1.047	1.052
Sugito	27.0	84	98	95	97	3.710	0.995	1.096	1.009
Matsubushi	16.4	109	101	100	100	6.078	1.049	1.054	1.056
Shouwa	24.4	85	98	95	97	4.098	1.008	1.096	1.021