CIRJE-F-982

Small Area Predictors with Dual Shrinkage of Means and Variances

Hiromasa Tamae
Graduate School of Economics, The University of Tokyo

Tatsuya Kubokawa
The University of Tokyo

July 2015
Small Area Predictors with Dual Shrinkage of Means and Variances

Hiromasa Tamae* and Tatsuya Kubokawa†

July 15, 2015

Abstract

The paper concerns small-area estimation in the Fay-Herriot type area-level model with random dispersions, which models the case that the sampling errors change from area to area. The resulting Bayes estimator shrinks both means and variances, but needs numerical computation to provide the estimates. In this paper, an approximated empirical Bayes (AEB) estimator with a closed form is suggested. The model parameters are estimated via the moment method, and the mean squared error of the AEB is estimated via the single parametric bootstrap method. The benchmarked estimator and a second-order unbiased estimator of the mean squared error are also derived.

Key words and phrases: Asymptotic approximation, benchmark, constrained Bayes, empirical Bayes, Fay-Herriot model, mean squared error, parametric bootstrap, random dispersion, second-order approximation, second-order unbiased estimate, small area estimation, variance modeling.

1 Introduction

Small area estimation (SAE) using linear mixed models has been extensively studied in the literature from both theoretical and applied points of view. For a good review and account on this topic, see Ghosh and Rao (1994), Pfeffermann (2002), Rao (2003) and Datta (2009). Of these, the Fay-Herriot model introduced by Fay and Herriot (1979) has been used as an area-level model in SAE.

*Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN, E-Mail: taisho.1603@gmail.com
†Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN. E-Mail: tatsuya@e.u-tokyo.ac.jp
Suppose that there are \( m \) small areas and that \( y_1, \ldots, y_m \) are direct estimates of small area means. The Fay-Herriot model is described as

\[
y_i \mid \xi_i \sim N(\xi_i, \sigma_i^2), \\
\xi_i \sim N(z_i^T \beta, \tau^2),
\]

where \( z_i \) is a vector of auxiliary variables and \( \beta \) is an unknown vector of regression coefficients. Although \( \sigma_i^2 \)'s are treated as known variances in the Fay-Herriot model, in practice, \( \sigma_i^2 \) are estimated quantities, and the resulting empirical Bayes (EB) estimators involve substantial estimation errors. To take this point into account, we suppose that statistics \( V_1, \ldots, V_m \) are available for estimating \( \sigma_i^2 \) and that \( V_i/\sigma_i^2 \) has a chi-square distribution with \( n_i \) degrees of freedom. Then, Wang and Fuller (2003) provided estimators of the mean squared error (MSE) of the empirical Bayes estimators. For such variance modeling approaches, see Arora and Lahiri (1997), You and Chapman (2006), Dass, Maiti, Ren and Sinha (2012), Jiang and Nguyen (2012). Also see Maiti, Ren and Sinha (2014) and the references therein.

In the Fay-Herriot models with heteroscedastic unknown variances, each variance \( \sigma_i^2 \) cannot be estimated consistently based on \( V_i \) when \( n_i \)'s are bounded. This leads to the inconsistency properties of estimation procedures, namely, the empirical Bayes estimator does not converge to the Bayes estimator, and the MSE of the empirical Bayes estimator cannot be estimated consistently. To fix this difficulty, Maiti, et al. (2014) suggested that \( \sigma_i^2 \) has an inverse gamma distribution. It is interesting to point out that the resulting empirical Bayes (EB) estimator of \( \xi_i \) shrinks both means and variances. Since the EB includes integration with respect to \( \sigma_i^2 \), however, the EB cannot be expressed in closed forms. Thus one needs numerical integration to provide values of the EB. Maiti, et al. (2014) derived a second-order unbiased estimator of the conditional mean squared error (cMSE) of the EB given \( (y_i, V_i) \). However, one needs heavy numerical computation to provide values of the estimator of cMSE. For unconditional MSE of the EB, no computational algorithm was provided in Maiti, et al. (2014), because the computation may be much harder.

In this paper, we consider to approximate the Bayes estimator in the Fay-Herriot random dispersion model given in Maiti, et al. (2014). Approximating the joint probability density function, we suggest the the approximated Bayes estimator

\[
\xi_i^{AB} = z_i^T \beta + \left( 1 - \frac{1}{1 + \tau^2 (n_i + 1 + \alpha) / (V_i + \gamma)} \right) (y_i - z_i^T \beta),
\]

where \( \alpha \) and \( \gamma \) are model parameters in the distribution of \( \sigma_i^2 \). Since

\[
\frac{V_i + \gamma}{n_i + 1 + \alpha} = \frac{n_i + 1}{n_i + 1 + \alpha} \frac{V_i}{n_i + 1} + \left( 1 - \frac{n_i + 1}{n_i + 1 + \alpha} \right) \frac{\gamma}{\alpha},
\]

the estimator \( \xi_i^{AB} \) is a dual shrinkage estimator with shrinking \( y_i \) towards \( z_i^T \beta \) and shrinking \( V_i/(n_i + 1) \) towards \( \gamma/\alpha \). This approximation is valid in the case of large \( n_i \), but we
want to use this closed-form estimator even for small \( n_i \). For the purpose, we need to evaluate the estimation error. Since \( \zeta_{i}^{AB} \) includes the model parameters \( \beta, \tau^2, \alpha \) and \( \gamma \), we estimate \( \beta \) with the generalized least squares estimator based on the approximated pdf and the other parameters \( \tau^2, \alpha \) and \( \gamma \) via the moment methods. We show the consistency of the suggested estimators for the model parameters. Plugging-in the consistent estimators in \( \zeta_{i}^{AB} \) yields the approximated empirical Bayes (AEB) estimator \( \hat{\zeta}_{i}^{AEB} \). The uncertainty of the AEB is measured via the the unconditional mean squared errors (MSE), and we obtain a second-order unbiased estimator of the MSE via the single parametric bootstrap method.

In this paper, we also treat the benchmark problem. A potential difficulty of the AEB estimators \( \hat{\zeta}_{i}^{AEB} \) for small areas is that the overall estimate for a larger geographical area, which is constructed by a (weighted) sum of \( \hat{\zeta}_{i}^{AEB} \), is not necessarily equal to the corresponding direct estimate like the overall sample mean. For instance, we consider the weighted mean \( \bar{y}_w = \sum_{j=1}^{m} w_j y_j \) for nonnegative constants \( w_j \)'s satisfying \( \sum_{j=1}^{m} w_j = 1 \). Then, we want to find predictors \( \delta_i \)'s which satisfy the benchmark constraint \( \sum_{j=1}^{m} w_j \delta_j = \bar{y}_w \). A solution of the benchmark problem is the constrained Bayes estimation suggested by Ghosh (1992) and Datta, Ghosh, Steorts and Maples (2011). Using this approach, we suggest the benchmarked predictor based on \( \hat{\zeta}_{i}^{AEB} \) given by

\[
\delta_i^{CAB} = \hat{\zeta}_i^{AEB} + \frac{w_i}{\sum_{j=1}^{m} w_j^2} \left\{ \bar{y}_w - \sum_{j=1}^{m} w_j \hat{\zeta}_j^{AEB} \right\}.
\]

A second-order unbiased estimator of the MSE of this constrained approximate Bayes estimator is derived.

The paper is organized as follows: A setup of the Fay-Herriot random dispersion model and the approximated Bayes estimator are given in Section 2. The estimators of the model parameters are also given there. In Section 3, the approximated empirical Bayes (AEB) estimator is evaluated in terms of the MSE, and the second-order unbiased estimator is suggested. The benchmark problem is discussed in Section 4. In Section 5, we investigate the performance of the proposed procedures through simulation and empirical studies. Concluding remarks are given in Section 6 and the technical proofs are given in the Appendix.

2 Area-level Model and Estimation of Model Parameters

2.1 Fay-Herriot random dispersion model and an approximated predictor

For \( m \) small areas, let \( \left( y_i, V_i/n_i \right) \) be the pair of mean estimate and variance estimate for the \( i \)-th small area, \( i = 1, \ldots, m \), where \( n_i \) is degrees of freedom. Suppose that there exist
where \( f \) are denoted by \( \mathbf{z}_i = (z_{i1}, \ldots, z_{ip}) \) with \( z_{i1} = 1 \). Then we consider the following heteroscedastic area-level model with random dispersions:

\[
\begin{align*}
\mathbf{y}_i | \mathbf{\xi}_i, \sigma_i^2 &\sim \mathcal{N}(\mathbf{\xi}_i, \sigma_i^2), \\
\mathbf{\xi}_i &\sim \mathcal{N}(\mathbf{z}_i^T \mathbf{\beta}, \tau^2), \\
\mathbf{V}_i/\sigma_i^2 &\sim \chi^2_n, \\
\sigma_i^{-2} &\sim \text{Ga}(\alpha/2, 2/\gamma),
\end{align*}
\]  

(2.1)

where \((y_1, V_1), \ldots, (y_m, V_m)\) are mutually independent. We call it the Fay-Herriot Random Dispersion model (hereafter, FHRD model). Here, \( \text{Ga}(\alpha/2, 2/\gamma) \) denotes a gamma distribution with mean \( \alpha/\gamma \) and variance \( 2\alpha/\gamma^2 \). The unknown parameters are denoted by \( \mathbf{\omega} = (\mathbf{\beta}^T, \tau^2, \alpha, \gamma)^T \) for \( \mathbf{\beta} = (\beta_1, \ldots, \beta_p)^T \).

Let \( \eta_i = 1/\sigma_i^2 \) and \( C_i = 1/[2\pi^{n_i/2}\Gamma(n_i/2)] \). The joint pdf of \((y_i, V_i, \mathbf{\xi}_i, \eta_i)\) is

\[
f_i(y_i, V_i, \mathbf{\xi}_i, \eta_i) = C_i \frac{(\gamma/2)^{n_i/2}}{\Gamma(\alpha/2)} V_i^{n_i/2-1} \eta_i^{(n_i+1+\alpha)/2-1} \times \exp \left[ -\frac{n_i}{2} ((y_i - \mathbf{\xi}_i)^2 + V_i + \gamma) - \frac{1}{2\tau^2} (\mathbf{\xi}_i - \mathbf{z}_i^T \mathbf{\beta})^2 \right].
\]  

(2.2)

It is noted that

\[
\eta_i(y_i - \mathbf{\xi}_i)^2 + \tau^{-2}(\mathbf{\xi}_i - \mathbf{z}_i^T \mathbf{\beta})^2 + \eta_i(V_i + \gamma)
\]

\[
= (\eta_i + \tau^{-2}) \{\mathbf{\xi}_i - \mathbf{\xi}_i^M(\eta_i)\}^2 + \eta_i \left\{ V_i + \gamma + \frac{1}{\tau^2 \eta_i + 1} (y_i - \mathbf{z}_i^T \mathbf{\beta}) \right\},
\]

where

\[
\mathbf{\xi}_i^M(\eta_i) = \mathbf{z}_i^T \mathbf{\beta} + \left( 1 - \frac{1}{\tau^2 \eta_i + 1} \right) (y_i - \mathbf{z}_i^T \mathbf{\beta}).
\]  

(2.3)

Then, the Bayes estimator of \( \mathbf{\xi}_i \) is described as

\[
\mathbf{\xi}_i^B = E[\mathbf{\xi}_i | y_i, V_i] = \mathbf{z}_i^T \mathbf{\beta} + \left( 1 - E \left[ \frac{1}{\tau^2 \eta_i + 1} | y_i, V_i \right] \right) (y_i - \mathbf{z}_i^T \mathbf{\beta}),
\]  

(2.4)

where

\[
E \left[ \frac{1}{\tau^2 \eta_i + 1} | y_i, V_i \right] = \frac{\int_0^\infty (\tau^2 \eta_i + 1)^{-1} f_i(y_i, V_i, \eta_i) d\eta_i}{\int_0^\infty f_i(y_i, V_i, \eta_i) d\eta_i},
\]  

(2.5)

for the marginal pdf of \( f_i(y_i, V_i, \eta_i) \) given by

\[
f_i(y_i, V_i, \eta_i) = C_i \frac{(\gamma/2)^{n_i/2}}{\Gamma(\alpha/2)} V_i^{n_i/2-1} \eta_i^{(n_i+1+\alpha)/2-1} \frac{\sqrt{2\pi}}{\sqrt{\tau^2 \eta_i + 1}} \times \exp \left[ -\frac{n_i}{2} V_i + \gamma + \frac{1}{\tau^2 \eta_i + 1} (y_i - \mathbf{z}_i^T \mathbf{\beta})^2 \right].
\]  

(2.6)
When $\sigma_i^2 = 1/\eta_i$ is fixed and unknown, it may be estimated with $V_i/n_i$. Then from (2.3), one gets the estimator

$$z_i^T \beta + \left(1 - \frac{1}{\tau^2(n_i/V_i) + 1}\right)(y_i - z_i^T \beta),$$

which is not very stable for small $n_i$ due to the estimation error in $V_i/n_i$. The Bayes estimator (2.4) can fix this undesirable property. However, we resort to numerical integration to obtain the Bayes estimator and the empirical Bayes estimator. It may be computationally harder to evaluate the mean squared error of the empirical Bayes estimator.

We want to suggest another estimator with a closed form. To this end, we begin by integrating out the joint density (2.2) with respect to $\eta_i$. Then the marginal pdf of $(y_i, V_i, \xi_i)$ is written as

$$h_i(y_i, V_i, \xi_i) = C_i \frac{(\gamma/2)^{n_i/2}}{\tau_i^{(\alpha/2)}} V_i^{n_i/2 - 1} \exp \left[-\frac{1}{2\tau_i^2} (\xi_i - z_i^T \beta)^2\right]$$

$$\times \Gamma\left(\frac{n_i + 1 + \alpha}{2}\right) \left(\frac{2}{\eta_i - \xi_i)^2 + V_i + \gamma}\right)^{(n_i + 1 + \alpha)/2}.$$

Based on the density $h_i(y_i, V_i, \xi_i)$, the Bayes estimator of $\xi_i$ is also expressed as $\xi_i^B = E[\xi_i | y_i, V_i]$. We here consider to approximate the marginal pdf $h_i(y_i, V_i, \xi_i)$. It is noted that

$$\left(\frac{2}{\eta_i - \xi_i)^2 + V_i + \gamma}\right)^{(n_i + 1 + \alpha)/2}$$

$$= \left(\frac{2}{V_i + \gamma}\right)^{(n_i + 1 + \alpha)/2} \exp \left[-\frac{n_i + 1 + \alpha}{2} \log \left(1 + \frac{(y_i - \xi_i)^2}{V_i + \gamma}\right)\right],$$

Then, the function $\log\{1 + (y_i - \xi_i)^2/(V_i + \gamma)\}$ is approximated as

$$\log\{1 + (y_i - \xi_i)^2/(V_i + \gamma)\} \approx (y_i - \xi_i)^2/(V_i + \gamma). \quad (2.7)$$

This approximation can be guaranteed when $n_i$ is large. However, we use this approximation for small $n_i$ as well, and derive estimators of the unknown parameters and predictors for $\xi_i$ based on this approximation.

Using this approximation, we can rewrite the pdf $h_i(y_i, V_i, \xi_i)$ as

$$h_i^*(y_i, V_i, \xi_i) = C_i \frac{(\gamma/2)^{n_i/2}}{\tau_i^{(\alpha/2)}} V_i^{n_i/2 - 1} \left(\frac{2}{V_i + \gamma}\right)^{(n_i + 1 + \alpha)/2} \Gamma\left(\frac{n_i + 1 + \alpha}{2}\right)$$

$$\times \exp \left[-\frac{1}{2\tau_i^2} (\xi_i - z_i^T \beta)^2 - \frac{A_i}{2} (y_i - \xi_i)^2\right],$$

for $A_i = (n_i + 1 + \alpha)/(V_i + \gamma)$. It is noted that $h_i^*(y_i, V_i, \xi_i)$ is not a pdff. Since

$$\frac{1}{\tau^2}(\xi_i - z_i^T \beta)^2 + A_i(y_i - \xi_i)^2 = \frac{1 + \tau_i^2 A_i}{\tau_i^2} (\xi_i - z_i^T \beta + \tau_i^2 A_i y_i)^2 + \frac{A_i}{1 + \tau_i^2 A_i} (y_i - z_i^T \beta)^2,$$
we get the approximated Bayes estimator of $\xi_i$ given by

$$\xi_i^{AB} = \xi_i^{AB}(\beta, \theta) = \frac{z_i^T \beta + \tau^2 A_i y_i}{1 + \tau^2 A_i} = z_i^T \beta + (1 - B_i)(y_i - z_i^T \beta),$$

(2.8)

where $\theta = (\tau^2, \alpha, \gamma)^T$ and

$$B_i = B_i(\theta, V_i) = \frac{1}{1 + \tau^2 A_i} = \frac{1}{1 + \tau^2(n_i + 1 + \alpha)/(V_i + \gamma)}.$$

It is noted that this is not the Bayes estimator, but the approximated Bayes estimator when the approximation (2.7) is valid. Since the approximated Bayes estimator has a simple and reasonable form, however, we shall use this estimator even if this approximation is not appropriate. The following proposition implies that the approximated Bayes estimator $\xi_i^{AB}$ has less shrinkage than the Bayes estimator $\xi_i^B$ given in (2.4). The proof is given in the Appendix.

**Proposition 2.1** The shrinkage function $B_i$ in $\xi_i^{AB}$ is less than the shrinkage function (2.5) in the Bayes estimator $\xi_i^B$, namely,

$$E\left[\frac{1}{\tau^2 \eta_i + 1} \mid y_i, V_i\right] \geq B_i = \frac{1}{1 + \tau^2(n_i + 1 + \alpha)/(V_i + \gamma)}.$$

### 2.2 Estimation of the model parameters

We now provide estimators of the model parameters $\beta$, $\tau^2$, $\alpha$ and $\gamma$.

**[1] Estimation of $\beta$.** Integrating out $h_i^*(y_i, V_i, \xi_i)$ with respect to $\xi_i$, we have

$$h_i^*(y_i, V_i) = C_i \frac{(\gamma/2)^{\alpha/2}}{\Gamma(\alpha/2)} V_i^{n_i/2 - 1} \left(\frac{2}{V_i + \gamma}\right)^{(n_i + 1 + \alpha)/2} \Gamma\left(\frac{n_i + 1 + \alpha}{2}\right) \times \sqrt{2\pi} \sqrt{B_i} \exp\left[-\frac{1}{2\tau^2}(1 - B_i)(y_i - z_i^T \beta)^2\right].$$

Let $\ell^* = \ell^*(\beta, \tau^2, \alpha, \gamma) = \sum_{i=1}^m \ell_i^*$ for $\ell_i^* = \log h_i^*(y_i, V_i)$. Since

$$2 \frac{\partial \ell^*}{\partial \beta} = \frac{1}{\tau^2} \sum_{i=1}^m (1 - B_i)(y_i - z_i^T \beta)z_i,$$

we get the estimator

$$\tilde{\beta} = \tilde{\beta}(\beta, \tau^2, \alpha, \gamma) = \left(\sum_{j=1}^m (1 - B_j)z_j z_j^T\right)^{-1} \sum_{j=1}^m (1 - B_j)z_j y_j,$$

(2.9)

which is the generalized least squares (GLS) estimator of $\beta$. 
2. Estimation of $\tau^2$. To estimate $\tau^2$, we consider the expectation $E[(y_i - z_i^T \beta)^2/(V_i + \gamma)]$. The conditional expectation of $(y_i - z_i^T \beta)^2$ given $V_i$ is decomposed as

$$E[(y_i - z_i^T \beta)^2 | V_i] = E[(y_i - \xi_i)^2 | V_i] + 2E[(y_i - \xi_i)(\xi_i - z_i^T \beta) | V_i] + E[(\xi_i - z_i^T \beta)^2 | V_i].$$

Since $E[(y_i - \xi_i)^2 | V_i, \eta_i, \xi_i] = 1/\eta_i$, $E[(y_i - \xi_i)(\xi_i - z_i^T \beta) | V_i, \eta_i, \xi_i] = 0$ and $E[(\xi_i - z_i^T \beta)^2 | V_i, \eta_i] = \tau^2$, it is seen that

$$E[(y_i - z_i^T \beta)^2 | V_i] = E[\eta_i^{-1} | V_i] + \tau^2.$$

The joint pdf of $(V_i, \eta_i)$ is

$$f_i(V_i, \eta_i) = \frac{\gamma^{\alpha/2} V_i^{n_i/2-1}}{\Gamma(n_i/2)\Gamma(\alpha/2)^2(n_i+\alpha)/2} \eta_i^{(n_i+\alpha)/2-1} e^{-(\eta_i/2)(V_i+\gamma)}, \quad (2.10)$$

so that the marginal pdf of $V_i$ is

$$f_i(V_i) = \frac{\Gamma((n_i + \alpha)/2)}{\Gamma(n_i/2)\Gamma(\alpha/2)^2} (V_i + \gamma)^{(n_i+\alpha)/2}. \quad (2.11)$$

and the conditional pdf of $\eta_i$ given $V_i$ is

$$\eta_i | V_i \sim Ga\left(\frac{n_i + \alpha}{2}, \frac{2}{V_i + \gamma}\right). \quad (2.12)$$

Thus, one gets

$$E[\eta_i^{-1} | V_i] = \frac{V_i + \gamma}{n_i + \alpha - 2},$$

which implies that

$$E[(y_i - z_i^T \beta)^2 | V_i] = \frac{V_i + \gamma}{n_i + \alpha - 2} + \tau^2. \quad (2.13)$$

Thus, from this equality, we consider the moment $E[(y_i - z_i^T \beta)^2/(V_i + \gamma)]$, which is

$$E\left[\frac{(y_i - z_i^T \beta)^2}{V_i + \gamma}\right] = \frac{1}{n_i + \alpha - 2} + E\left[\frac{\tau^2}{V_i + \gamma}\right].$$

To calculate the moments of $V_i$ from the marginal pdf (2.11), the following equality is useful: In general, for real numbers $\ell$ and $k$, it can be shown that

$$E\left[\frac{V_i^\ell}{(V_i + \gamma)^k}\right] = \frac{\Gamma((n_i + \alpha)/2) \Gamma(\alpha/2)^2(n_i+\alpha)/2 + \ell)}{\Gamma((n_i + \alpha)/2 + 2 + k) \Gamma(n_i/2) \Gamma(\alpha/2) \Gamma(\alpha/2) \gamma^\ell - k}. \quad (2.14)$$

For $\ell = 0$ and $k = 1$, we have $E[1/(V_i + \gamma)] = \alpha/\{\gamma(n_i + \alpha)\}$, so that

$$E\left[\frac{(y_i - z_i^T \beta)^2}{V_i + \gamma}\right] = \frac{1}{n_i + \alpha - 2} + \frac{\alpha/\gamma}{n_i + \alpha} \tau^2.$$
When \( \alpha \) and \( \gamma \) are known and \( \beta \) is estimated by the ordinary least squares (OLS) estimator
\[
\hat{\beta}_{OLS} = \left( \sum_{j=1}^{m} z_j z_j^T \right)^{-1} \sum_{j=1}^{m} z_j y_j,
\]
this gives us the estimator
\[
\hat{\tau}^2 = \left( \sum_{i=1}^{m} \frac{\alpha/\gamma}{n_i + \alpha} \right)^{-1} \sum_{i=1}^{m} \left\{ \frac{(y_i - \bar{z}_i \hat{\beta}_{OLS})}{V_i + \gamma} - \frac{1}{n_i + \alpha - 2} \right\}.
\] (2.15)

[3] Estimation of \( \alpha \). Concerning the estimation of \( \alpha \), we concentrate on the marginal pdf (2.11) of \( V_i \). Since
\[
\log f_i(V_i) = \log \Gamma \left( \frac{n_i + \alpha}{2} \right) \quad \text{and} \quad \log \Gamma \left( \frac{\alpha}{2} \right) \quad \text{and} \quad \frac{n_i}{2} \log \Gamma \left( \frac{\alpha}{2} \right),
\]
we have
\[
2 \frac{\partial}{\partial \alpha} \log f_i(V_i) = \psi \left( \frac{n_i}{2} + \frac{\alpha}{2} \right) - \psi \left( \frac{\alpha}{2} \right) + \log \gamma - \log(V_i + \gamma),
\]
where \( \psi(\cdot) \) is the digamma function given by \( \psi(x) = \Gamma'(x)/\Gamma(x) \). Since \( E[\partial f_i(V_i)/\partial \alpha] = 0 \), one gets
\[
E[\log(V_i + \gamma)] = \psi \left( \frac{n_i}{2} + \frac{\alpha}{2} \right) - \psi \left( \frac{\alpha}{2} \right) + \log \gamma.
\] (2.16)

We here note the following equality. For real numbers \( \ell \) and \( k \), it can be shown that
\[
E \left[ \frac{V_i^\ell}{(V_i + \gamma)^k} \log(V_i + \gamma) \right] = \frac{\Gamma((n_i + \alpha)/2) \Gamma(n_i/2 + \ell) \Gamma(\alpha/2 + k - \ell)}{\Gamma((n_i + \alpha)/2 + k) \Gamma(n_i/2) \Gamma(\alpha/2)} \gamma^{\ell-k} \times \left\{ \psi \left( \frac{n_i + \alpha}{2} + k \right) - \psi \left( \frac{\alpha}{2} + k - \ell \right) + \log \gamma \right\}.
\] (2.17)

For \( \ell = 1 \) and \( k = 1 \), we have
\[
E \left[ \frac{V_i}{V_i + \gamma} \log(V_i + \gamma) \right] = \frac{n_i}{n_i + \alpha} \left\{ \psi \left( \frac{n_i + \alpha}{2} + 1 \right) - \psi \left( \frac{\alpha}{2} \right) + \log \gamma \right\}.
\]
Since the digamma function has the property that \( \psi(x + 1) = \psi(x) + 1/x \), it follows from (2.16) that
\[
E \left[ \frac{V_i}{V_i + \gamma} \log(V_i + \gamma) \right] = \frac{n_i}{n_i + \alpha} \left\{ E[\log(V_i + \gamma)] + \frac{2}{n_i + \alpha} \right\}.
\] (2.18)

This can be rewritten as
\[
\alpha^2 E \left[ \frac{V_i}{V_i + \gamma} \log(V_i + \gamma) \right] + \alpha E \left[ n_i \frac{V_i - \gamma}{V_i + \gamma} \log(V_i + \gamma) \right] - n_i^2 E \left[ \gamma \frac{\gamma}{V_i + \gamma} \log(V_i + \gamma) \right] - 2n_i = 0,
\]
which yields an estimator of $\alpha$. In fact, we can suggest the estimator as the solution of the quadratic equation
\[
\alpha^2 \sum_{i=1}^{m} \frac{V_i}{V_i + \gamma} \log(V_i + \gamma) + \alpha \sum_{i=1}^{m} n_i \frac{V_i - \gamma}{V_i + \gamma} \log(V_i + \gamma)
\]
\[
- \sum_{i=1}^{m} n_i \left\{ \frac{n_i \gamma}{V_i + \gamma} \log(V_i + \gamma) + 2 \right\} = 0.
\]
(2.19)

[4] Estimation of $\gamma$. Concerning the estimation of $\gamma$, from (2.14), it follows that
\[
E\left[ \frac{V_i}{V_i + \gamma} \right] = \frac{n_i}{n_i + \tilde{\alpha}}.
\]
Thus, one gets the estimator of $\gamma$ as the solution of the equation
\[
\sum_{i=1}^{m} \frac{V_i}{V_i + \gamma} = \sum_{i=1}^{m} \frac{n_i}{n_i + \tilde{\alpha}}.
\]
(2.20)

3 Evaluation of Uncertainty of Prediction

Substituting the estimators of $\beta$, $\tau^2$, $\alpha$ and $\gamma$ into (2.8), we get the predictor
\[
\hat{\xi}_i^{AB} = \xi_i^{AB}(\hat{\beta}, \hat{\tau^2}, \hat{\alpha}, \hat{\gamma}) = z_i^T \hat{\beta} + (1 - \hat{B}_i)(y_i - z_i^T \hat{\beta}),
\]
(3.1)
where
\[
\hat{B}_i = \frac{1}{1 + \hat{\tau}^2(n_i + 1 + \hat{\alpha})/(V_i + \hat{\gamma})}.
\]
(3.2)
\[
\hat{\beta} = \left( \sum_{j=1}^{m} (1 - \hat{B}_j)z_jz_j^T \right)^{-1} \sum_{j=1}^{m} (1 - \hat{B}_j)z_jy_j.
\]
(3.3)
We call it the approximated empirical Bayes estimator. It is noted that the term $(V_i + \hat{\gamma})/(n_i + 1 + \hat{\alpha})$ in $\hat{B}_i$ is expressed as
\[
\frac{V_i + \hat{\gamma}}{n_i + 1 + \hat{\alpha}} = \frac{n_i + 1}{n_i + 1 + \tilde{\alpha}} \frac{V_i}{n_i + 1} + \left( 1 - \frac{n_i + 1}{n_i + 1 + \tilde{\alpha}} \right) \frac{\hat{\gamma}}{\tilde{\alpha}},
\]
which shrinks $V_i/(n_i + 1)$ towards the target $\hat{\gamma}/\tilde{\alpha}$. Thus, the predictor $\hat{\xi}_i^{AB}$ is a double shrinkage procedure such that $y_i$ and $V_i/(n_i + 1)$ are shrunk towards $z_i^T \hat{\beta}$ and $\hat{\gamma}/\tilde{\alpha}$, respectively. In this section, we derive a second-order unbiased estimator of the mean squared error (MSE) of $\hat{\xi}_i^{AB}$.

We begin by rewriting the predictor as
\[
\hat{\xi}_i^{AB} = \{(1 - B_i)y_i + B_iz_i^T \beta\} - \{(\hat{B}_i - B_i)(y_i - z_i^T \beta) - \hat{B}_iz_i^T (\hat{\beta} - \beta)\}.
\]
Thus, the MSE of $\hat{\xi}^{AEB}_i$ is decomposed as

$$MSE(\hat{\xi}^{AEB}_i) = E[(\hat{\xi}_i - \xi_i)^2] = g_1 + g_2 - 2g_3,$$  

(3.4)

where

$$g_1 = E[((1 - B_i)y_i + B_i z_i^T \beta - \xi_i)^2],$$

$$g_2 = E[((\hat{B}_i - B_i)(y_i - z_i^T \beta) - \hat{B}_i z_i^T (\hat{\beta} - \beta))^2],$$

$$g_3 = E[((1 - B_i)y_i + B_i z_i^T \beta - \xi_i)((\hat{B}_i - B_i)(y_i - z_i^T \beta) - \hat{B}_i z_i^T (\hat{\beta} - \beta))].$$

(3.5)

To evaluate $g_1$, $g_2$ and $g_3$, we use the following theorem under the assumption (A):

**Assumption (A)**

(A1) There exist $p$ and $\pi$ such that $p \leq n_i \leq \pi$ for $i = 1, \ldots, m$. The dimension $p$ is bounded.

(A2) The matrix $m^{-1} \sum_{i=1}^m z_i z_i^T$ converges to a positive definite matrix.

**Theorem 3.1** Assume the condition (A) and $n_i + \alpha > 4$ for $i = 1, \ldots, m$. Then, $E[\hat{\omega} - \omega] = O(m^{-1})$ and $E[(\hat{\omega} - \omega)(\hat{\omega} - \omega)^T] = O(m^{-1})$. Also, the conditional moments given $y_i, V_i$ satisfy that $E[\hat{\omega} - \omega \mid y_i, V_i] = E[\hat{\omega} - \omega] + o_p(m^{-1})$ and $E[(\hat{\omega} - \omega)(\hat{\omega} - \omega)^T \mid y_i, V_i] = E[(\hat{\omega} - \omega)(\hat{\omega} - \omega)^T] + o_p(m^{-1})$.

We first evaluate $g_1$. Since $(1 - B_i)y_i + B_i z_i^T \beta - \xi_i = (1 - B_i)(y_i - z_i^T \beta) - (\xi_i - z_i^T \beta)$ and $B_i$ is a function of $V_i$, it is seen that

$$g_1 = E[(1 - B_i)^2(y_i - z_i^T \beta)^2] + E[(\xi_i - z_i^T \beta)^2] - 2E[(1 - B_i)(y_i - z_i^T \beta)(\xi_i - z_i^T \beta)]$$

$$= E[(1 - B_i)^2(\sigma_i^2 + \tau^2) + \tau^2 - 2(1 - B_i)\tau^2].$$

Note that $E[\sigma_i^2 \mid V_i] = E[1/\eta_i \mid V_i] = (V_i + \gamma)/(n_i - 2 + \alpha)$. Thus, one gets $g_1 = E[G(\theta, V_i)]$, where

$$G(\theta, V_i) = \frac{V_i + \gamma}{n_i - 2 + \alpha}(1 - B_i)^2 + \tau^2 B_i^2,$$  

(3.6)

which is of order $O_p(1)$. We here rewrite $g_1$ as

$$g_1 = E[G(\hat{\theta}, V_i)] - E[G(\theta, V_i) - G(\theta, V_i)] = g_{11} - g_{12}. \quad \text{(say)}$$

An exact unbiased estimator of $g_{11}$ is $G(\hat{\theta}, V_i)$. Concerning $g_{12}$, the Taylor series expansion give us the approximation as

$$g_{12} = E[(\hat{\theta} - \theta)^T \frac{\partial}{\partial \theta} G(\theta, V_i)] + \frac{1}{2} E[(\hat{\theta} - \theta)^T (\frac{\partial^2}{\partial \theta \partial \theta^T} G(\theta, V_i))(\hat{\theta} - \theta)] + O(m^{-3/2}).$$
Since $\hat{\omega} - \omega = O_p(m^{-1/2})$ from Theorem 3.1, it is clear that the second term in $g_{12}$ is of order $O(m^{-1})$. For the first term, it follows from Theorem 3.1 that
\[
E\left[ (\hat{\theta} - \theta)^T \frac{\partial}{\partial \theta} G(\theta, V_i) \mid y_i, V_i \right] = E[(\hat{\theta} - \theta)^T \mid y_i, V_i] \frac{\partial}{\partial \theta} G(\theta, V_i) \\
= E[(\hat{\theta} - \theta)^T] \frac{\partial}{\partial \theta} G(\theta, V_i) + o_p(m^{-1}),
\]
which is of order $O_p(m^{-1})$. This shows that $g_{12} = O(m^{-1})$.

For $g_2$, it is clear that $g_2 = O(m^{-1})$. For $g_3$, it is noted that
\[
g_3 = E\{(1 - B_i)y_i + B_i z_i^T \beta - E[\xi_i \mid y_i, V_i]\\} \{E[\hat{B}_i - B_i \mid y_i, V_i](y_i - z_i^T \beta) - B_i z_i^T (\hat{\beta} - \beta)\} \\
- E\{(1 - B_i)y_i + B_i z_i^T \beta - E[\xi_i \mid y_i, V_i] (\hat{B}_i - B_i) z_i^T (\hat{\beta} - \beta)\}.
\]
The second term is of order $O(m^{-1})$. For the first term,
\[
E\{(1 - B_i)y_i + B_i z_i^T \beta - E[\xi_i \mid y_i, V_i]\\} \{E[\hat{B}_i - B_i \mid y_i, V_i](y_i - z_i^T \beta) - B_i z_i^T E[\hat{\beta} - \beta \mid y_i, V_i]\},
\]
which is approximated as
\[
E\{(1 - B_i)y_i + B_i z_i^T \beta - E[\xi_i \mid y_i, V_i]\\} \{E[\hat{B}_i - B_i \mid y_i - z_i^T \beta] - B_i z_i^T E[\hat{\beta} - \beta]\} + o(m^{-1}).
\]
This shows that $g_3 = O(m^{-1})$. Hence, we get the following proposition.

**Proposition 3.1** Assume the condition (A) and $n_i + \alpha > 4$ for $i = 1, \ldots, m$. Then, the MSE of the predictor $\hat{\xi}_i^{AEB}$ is decomposed as
\[
MSE(\hat{\xi}_i^{AEB}) = g_{11} + \{-g_{12} + 2g_2 - 2g_3\},
\]
where $g_{11} = O(1)$, $g_{12} = O(m^{-1})$, $g_2 = O(m^{-1})$ and $g_3 = O(m^{-1})$.

We next estimate the MSE of $\hat{\xi}_i^{AEB}$. An exact unbiased estimator of $g_{11}$ is given by
\[
\hat{g}_{11} = G(\hat{\theta}, V_i) = \frac{V_i + \hat{\gamma}}{n_i - 2 + \hat{\alpha}} (1 - \hat{B}_i)^2 + \hat{\tau}^2 \hat{B}_i^2.
\]
To provide second-order unbiased estimators of $g_{12}$, $g_2$ and $g_3$, we use the parametric bootstrap method. Let $(y_i^*, V_i^*)$, $i = 1, \ldots, m$, be a bootstrap sample generated from the model:
\[
y_i^* \mid \xi_i^*, \eta_i^* \sim N(\xi_i^*, 1/\eta_i^*), \\
\xi_i^* \sim N(z_i^T \hat{\beta}, \hat{\tau}^2), \\
V_i^* \mid \eta_i^* \sim \chi^2_{n_i}, \\
\eta_i^* \sim Ga(\hat{\alpha}/2, 2/\hat{\gamma}), \quad (3.7)
\]
where $\hat{\beta}$, $\hat{\tau}^2$, $\hat{\alpha}$ and $\hat{\gamma}$ are estimators constructed from the original model (2.1). The bootstrap estimators $\hat{\beta}^*, \hat{\tau}^*$, $\hat{\alpha}^*$ and $\hat{\gamma}^*$ are calculated via the same manner as in $\hat{\beta}$, $\hat{\tau}^2$, $\hat{\alpha}$
and \( \hat{\gamma} \) except that the bootstrap estimators are calculated based on \((y_i^*, V_i^*)\)'s instead of \((y_i, V_i)\)'s. Then we can estimate \( g_{12}, g_2 \) and \( g_3 \) with
\[
\begin{align*}
g_{12}^* &= E^*[G(\hat{\theta}^*, V_i^*) - G(\hat{\theta}, V_i^*)], \\
g_2^* &= E^*[\{(\hat{B}_i^* - B_i^*)(y_i^* - z_i^T \beta) - \hat{B}_i^* z_i^T (\beta^* - \hat{\beta})\}^2], \\
g_3^* &= E^*[\{(1 - B_i^*)y_i^* + B_i^* z_i^T \beta - \xi_i^*\}(\hat{B}_i^* - B_i^*) (y_i^* - z_i^T \beta) - \hat{B}_i^* z_i^T (\beta^* - \hat{\beta})]],
\end{align*}
\]
where \( \hat{\theta}^* = (\hat{\tau}^*, \hat{\alpha}^*, \hat{\gamma}^*)^T \), \( B_i^* = B_i(\hat{\theta}, V_i^*) \) and \( \hat{B}_i^* = B_i(\hat{\theta}^*, V_i^*) \). It can be seen that these are second-order unbiased estimators.

**Proposition 3.2** Assume the condition (A) and \( n_i + \alpha > 4 \) for \( i = 1, \ldots, m \). Then, a second-order unbiased estimator of the MSE of \( \hat{\xi}_i^{AEB} \) is
\[
mse(\hat{\xi}_i^{AEB}) = \hat{g}_{11} + \{-g_{12}' + g_2' - 2g_3'\},
\]
where \( E[\hat{g}_{11}] = g_{11}, E[g_{12}'] = g_{12} + o(m^{-1}), E[g_2'] = g_2 + o(m^{-1}) \) and \( E[g_3'] = g_3 + o(m^{-1}) \).

## 4 Benchmark Prediction

In this section, we consider the benchmark problem which imposes a constraint on predictors for small areas. The benchmarked predictors are derived and an approximated unbiased estimator of their MSE is provided.

Although the predictors \( \hat{\xi}_i \) in (3.1) give reliable estimates for \( \xi_i \) by borrowing strength from the surrounding areas, we are faced with a potential difficulty of the predictor. That is, the overall estimate for a larger geographical area, which is constructed by a (weighted) sum of \( \hat{\xi}_i \), is not necessarily equal to the corresponding direct estimate like the overall sample mean. To describe it specifically, let \( w_j \)'s be nonnegative constants such that \( \sum_{j=1}^m w_j = 1 \). Suppose that the mean of the total areas is estimated by the weighted sum of \( y_j \)'s, \( \sum_{j=1}^m w_j y_j \). Then, the benchmark problem is described as an issue of finding estimators \( \hat{\delta}_j \) such that
\[
\sum_{j=1}^m w_j \delta_j = \sum_{j=1}^m w_j y_j = \bar{y}_w. \tag{4.1}
\]
A solution of the benchmark problem is the constrained Bayes estimation suggested by Ghosh (1992) and Datta, Ghosh, Steorts and Maples (2011), who considered the minimization of \( \sum_{i=1}^m E[(\delta_i - \xi_i)^2 \mid Data] \) under the constraint (4.1). Using the Lagrange multiplier method, one gets the constrained Bayes estimator
\[
\delta_i^{CB} = E[\xi_i \mid y_i, V_i] + \frac{w_i}{\sum_{j=1}^m w_j} \left\{ \bar{y}_w - \sum_{j=1}^m w_j E[\xi_j \mid y_j, V_j] \right\}.
\]

12
Since the Bayes estimator $E[\xi_i \mid y_i, V_i]$ cannot be expressed in a closed form, we replace it with the approximated empirical Bayes estimator $\hat{\xi}_i$ given in (3.1). The resulting benchmarked predictor is

$$\delta_i^{CAB} = \hat{\xi}_i^{AEB} + \frac{w_i}{\sum_{j=1}^{m} w_j^2} \left\{ \bar{y}_w - \sum_{j=1}^{m} w_j \hat{\xi}_j^{AEB} \right\},$$

(4.2)

which is here called the constrained approximate Bayes estimator.

To evaluate the uncertainty of $\delta_i^{CAB}$, we derive a second-order unbiased estimator of the MSE. The MSE of $\delta_i^{CAB}$ is decomposed as

$$E[(\delta_i^{CAB} - \xi_i)^2] = E[(\hat{\xi}_i^{AEB} - \xi_i)^2] + \frac{w_i^2}{(\sum_{j=1}^{m} w_j^2)^2} E\left[ (\bar{y}_w - \sum_{j=1}^{m} w_j \hat{\xi}_j^{AEB})^2 \right] + 2 \frac{w_i}{\sum_{j=1}^{m} w_j^2} J,$$

where

$$J = E\left[ (\hat{\xi}_i^{AEB} - \xi_i) \left( \bar{y}_w - \sum_{j=1}^{m} w_j \hat{\xi}_j^{AEB} \right) \right].$$

The second-order unbiased estimator of the first term $E[(\hat{\xi}_i^{AEB} - \xi_i)^2]$ is given in Proposition 3.2. Clearly, an exact unbiased estimator of $E[(\bar{y}_w - \sum_{j=1}^{m} w_j \hat{\xi}_j^{AEB})^2]$ is $(\bar{y}_w - \sum_{j=1}^{m} w_j \hat{\xi}_j^{AEB})^2$. In Theorem 4.1 given below, we verify that $J = O(1)$. Then, we can estimate $J$ based on the bootstrap sample given in (3.7) as

$$J^* = E^* \left[ (\hat{\xi}_i^{AEB*} - \xi_i^*) \left( \bar{y}_w^* - \sum_{j=1}^{m} w_j \hat{\xi}_j^{AEB*} \right) \right],$$

which satisfies that $E[J^*] = J + o(1)$.

**Theorem 4.1** Assume the condition (A) and $n_i + \alpha > 4$ for $i = 1, \ldots, m$. Also assume that $\sum_{j=1}^{m} w_j^2/m$ converges to a non-zero constant. Then, $J = O(1)$ and a second-order unbiased estimator of the MSE of $\delta_i^{CAB}$ is

$$\text{mse}(\delta_i^{CAB}) = \text{mse}(\hat{\xi}_i^{AEB}) + \frac{w_i^2}{(\sum_{j=1}^{m} w_j^2)^2} \left[ (\bar{y}_w - \sum_{j=1}^{m} w_j \hat{\xi}_j^{AEB})^2 \right] + 2 \frac{w_i}{\sum_{j=1}^{m} w_j^2} J^*,$$

(4.3)

where $\text{mse}(\hat{\xi}_i^{AEB})$ is given in (3.9). That is, $E[\text{mse}(\delta_i^{CAB})] = E[(\delta_i^{CAB} - \xi_i)^2] + o(m^{-1})$.

The proof of Theorem 4.1 is given in the Appendix.

## 5 Numerical and Empirical Studies

In this section, we investigate performances of the procedures suggested in the previous sections through numerical and empirical studies.
5.1 Simulation study

Here we investigate finite sample performances of the estimators in the Fay-Herriot random dispersion (FHRD) model and the second-order unbiased estimators for the unconditional MSEs by the Monte Carlo simulation. Comparing the performances of the approximated Bayes estimator $\xi_{AB}^i$ given in (2.8) with those of the Bayes estimator $\xi_i^B$ given in (2.4), we check goodness of the approximation we applied.

We conduct simulation experiments as we specified true model, so simulation data is generated by FHRD model (2.1). Throughout the simulations, the true value of $\beta$ and $\gamma$ are $\beta = 10$ and $\gamma = 1$. For each of $m, \tau^2, \alpha$, we examined two cases; $m = 30$ or $60$, $\tau^2 = 1$ or $4$ and $\alpha = 1$ or $4$. For simplicity, we set $z_i = 1$, $p = 1$ and $n_i = 10$ for all areas and cases. Thus, there are eight cases of simulations for the variety of $m$, $\alpha$ and $\tau^2$.

We first compute numerical values of MSE of the Bayes and approximate Bayes estimators $\xi_i^B$ and $\xi_{AB}^i$ with

$$\text{MSE}_i = \frac{1}{K} \sum_{k=1}^{K} (\hat{\xi}_i^{(k)} - \xi_i^{(k)})^2$$

for $K = 5,000$ where $\hat{\xi}_i^{(k)}$ and $\xi_i^{(k)}$ are the estimator and the true value of $\xi_i$ in the $k$-th replication for $k = 1, \ldots, K$. To investigate the loss which arises from approximation (2.7), we compare the approximated Bayes estimator $\xi_{AB}^i$ with the Bayes estimator $\xi_i^B$ in terms of biases and true MSEs under known model parameters. The results of the simulation are reported in Table 1.

Table 1: Biases and square roots of MSE of $\xi_{AB}^i$ and $\xi_i^B$ in the FHRD model under known model parameters

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\alpha, \tau^2$</th>
<th>Bias (m=30)</th>
<th>SRMSE (m=30)</th>
<th>Bias (m=60)</th>
<th>SRMSE (m=60)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi_i^B$</td>
<td>1, 1</td>
<td>0.002</td>
<td>0.818</td>
<td>-0.002</td>
<td>0.822</td>
</tr>
<tr>
<td></td>
<td>1, 4</td>
<td>0.003</td>
<td>1.347</td>
<td>-0.005</td>
<td>1.346</td>
</tr>
<tr>
<td></td>
<td>4, 1</td>
<td>-0.005</td>
<td>0.528</td>
<td>0.000</td>
<td>0.528</td>
</tr>
<tr>
<td></td>
<td>4, 4</td>
<td>-0.005</td>
<td>0.629</td>
<td>-0.002</td>
<td>0.629</td>
</tr>
<tr>
<td>$\xi_{AB}^i$</td>
<td>1, 1</td>
<td>0.002</td>
<td>0.824</td>
<td>-0.002</td>
<td>0.828</td>
</tr>
<tr>
<td></td>
<td>1, 4</td>
<td>0.003</td>
<td>1.355</td>
<td>-0.005</td>
<td>1.355</td>
</tr>
<tr>
<td></td>
<td>4, 1</td>
<td>-0.006</td>
<td>0.530</td>
<td>0.001</td>
<td>0.530</td>
</tr>
<tr>
<td></td>
<td>4, 4</td>
<td>-0.005</td>
<td>0.631</td>
<td>-0.002</td>
<td>0.629</td>
</tr>
</tbody>
</table>

It is observed from Table 1 that the difference between $\xi_i^B$ and $\xi_{AB}^i$ is tiny. Even though $\xi_{AB}^i$ is dominated by $\xi_i^B$ as expected, the difference of the biases is little except the case $(\alpha, \tau^2) = (4,1)$. Moreover, the largest difference between the two SRMSEs is
0.009 for \( m = 60 \) and \((\alpha, \tau^2) = (1, 4)\). Thus, the approximation little affects the bias and MSE.

We next investigate finite sample performances of the estimators for the model parameters. In particular, it is worth remarking the estimation of \( \alpha \). Initially, we used the maximum likelihood estimator of \( \alpha \) from the marginal likelihood as Maiti et al. (2014). The MLE can be obtained by solving the equation based on the digamma functions. However, the numerical solutions for the MLE yield large variability. To avoid such an instable performance of the MLE, in this paper, we suggest the new consistent estimator given in (2.19). The performances of the suggested estimators for \( \beta, \tau^2, \alpha \) and \( \gamma \) are reported in Table 2, where means and standard deviations via simulation with 1,000 replications are given. Table 2 shows that the estimator of \( \beta \) is almost unbiased and has small standard deviation. For other estimators, both biases and standard deviations are moderated as \( m \) increases. Especially, the suggested estimator (2.19) of \( \alpha \) provides stable estimates and a good performance.

Table 2: Results of the estimation for model parameters \( \beta, \tau^2, \alpha \) and \( \gamma \) in the FHRD model for \( \beta = 10 \) and \( \gamma = 1 \) (Standard deviations are shown in the parentheses)

<table>
<thead>
<tr>
<th>( m )</th>
<th>( \alpha, \tau^2 )</th>
<th>( \hat{\beta} )</th>
<th>( \hat{\tau^2} )</th>
<th>( \hat{\alpha} )</th>
<th>( \hat{\gamma} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 1</td>
<td>10.001(0.331)</td>
<td>0.912(0.658)</td>
<td>1.041(0.160)</td>
<td>1.092(0.262)</td>
<td></td>
</tr>
<tr>
<td>1, 4</td>
<td>9.997(0.495)</td>
<td>3.658(1.834)</td>
<td>1.038(0.157)</td>
<td>1.092(0.269)</td>
<td></td>
</tr>
<tr>
<td>4, 1</td>
<td>10.000(0.217)</td>
<td>0.950(0.361)</td>
<td>4.067(0.538)</td>
<td>1.013(0.085)</td>
<td></td>
</tr>
<tr>
<td>4, 4</td>
<td>10.006(0.384)</td>
<td>3.876(1.235)</td>
<td>4.063(0.543)</td>
<td>1.015(0.085)</td>
<td></td>
</tr>
<tr>
<td>1, 1</td>
<td>9.999(0.226)</td>
<td>0.944(0.483)</td>
<td>1.135(0.189)</td>
<td>1.203(0.346)</td>
<td></td>
</tr>
<tr>
<td>1, 4</td>
<td>10.006(0.345)</td>
<td>3.883(1.321)</td>
<td>1.018(0.114)</td>
<td>1.036(0.167)</td>
<td></td>
</tr>
<tr>
<td>4, 1</td>
<td>10.000(0.148)</td>
<td>0.977(0.254)</td>
<td>4.029(0.377)</td>
<td>1.006(0.061)</td>
<td></td>
</tr>
<tr>
<td>4, 4</td>
<td>9.998(0.268)</td>
<td>3.928(0.880)</td>
<td>4.036(0.383)</td>
<td>1.005(0.062)</td>
<td></td>
</tr>
</tbody>
</table>

Finally, we compare the second-order unbiased estimator of the MSE of \( \hat{\xi}^{AEB} \) with the true MSE. Concerning the MSE, the true value is calculated via simulation with \( R = 5,000 \) replications as 3.4. Then, the mean values of the estimator for the MSE and their Percentage Relative Bias (RB) are calculated based on \( T = 1,000 \) simulation runs with each 1,000 bootstrap samples, where RB is defined by

\[
\text{RB}_i = 100 \frac{T^{-1} \sum_{t=1}^{T} \overline{\text{MSE}}_i^{(t)} - \text{MSE}_i}{\text{MSE}_i},
\]

for the MSE estimate \( \overline{\text{MSE}}_i^{(t)} \) in the \( t \)-th replication for \( t = 1, \ldots, T \). Note that both true MSE and its estimator are calculated based on our estimates of the model parameters.
Table 3 reports values of the true MSE, the second-order unbiased estimator \( \hat{MSE} \) and the corresponding Percentage Relative Biases. It is observed that the MSE estimates \( \hat{MSE} \) are close to the true values of the MSE, and their relative bias are small for both \( m = 30 \) and \( 60 \). Thus, the second-order unbiased estimator of the MSE performs well as an estimator of the MSE of \( \hat{\xi}_{i}^{AEB} \) in the FHRD model.

<table>
<thead>
<tr>
<th>Size</th>
<th>( \alpha, \tau^{2} )</th>
<th>MSE</th>
<th>( \hat{MSE} )</th>
<th>RB(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m = 30 )</td>
<td>1, 1</td>
<td>0.996</td>
<td>0.952</td>
<td>-4.379</td>
</tr>
<tr>
<td></td>
<td>4, 1</td>
<td>0.312</td>
<td>0.323</td>
<td>3.580</td>
</tr>
<tr>
<td></td>
<td>4, 4</td>
<td>0.416</td>
<td>0.414</td>
<td>-0.443</td>
</tr>
<tr>
<td>( m = 60 )</td>
<td>1, 1</td>
<td>0.835</td>
<td>0.833</td>
<td>-0.233</td>
</tr>
<tr>
<td></td>
<td>4, 1</td>
<td>0.294</td>
<td>0.298</td>
<td>1.284</td>
</tr>
<tr>
<td></td>
<td>4, 4</td>
<td>0.399</td>
<td>0.405</td>
<td>1.424</td>
</tr>
</tbody>
</table>

5.2 Illustrative examples

We apply the approximated empirical Bayes estimator and the estimator of the MSE to the data in the Survey of Family Income and Expenditure (SFIE) in Japan.

In this study, we use the data of the spending items ‘Education’ and ‘Health’ in the survey in 2014. For the spending item ‘Education’, the annual average spending (scaled by 1,000 Yen) at each capital city of 47 prefectures in Japan is denoted by \( y_{i} \) for \( i = 1, \ldots, 47 \), and each variance \( V_{i} \) is calculated based on data of the spending ‘Education’ at the same city in the past consecutive eight years. Although the annual average spendings in SFIE are reported every year, the sample sizes are around 50 for most prefectures. We apply the same manner to the spending item ‘Health’ to create \( y_{i} \) and \( V_{i} \). The data of the item ‘Education’ have high variability, but those of the item ‘Health’ have relatively lower one.

In addition to the SFIE data, we can use data in the National Survey of Family Income and Expenditure (NSFIE) for 47 prefectures. Since NSFIE is based on much larger sample than SFIE, the annual average spendings in NSEDI are more reliable, but this survey has been implemented every five years. In this study, we use the data of the spending items ‘Education’ and ‘Health’ of NSFIE in 2009 as covariates \( z_{i} \) for \( i = 1, \ldots, 47 \). Thus, we apply the FHRD model (2.1) to these examples, where \( z_{i}^{T}\beta = \beta_{0} + z_{i}\beta_{1} \).

We calculated the predicted values of \( \hat{\xi}_{i}^{AEB} \) and \( \hat{\xi}_{i}^{EB} \) and the estimates of the MSE of \( \hat{\xi}_{i}^{AEB} \) with the estimates \( V_{i} \). These values in seven prefectures around Tokyo are reported
in Tables 4 and 5 for the ‘Education’ and ‘Health’ data. For the ‘Education’ data, the estimates of the model parameters are \( \hat{\beta}_0 = 15.711, \hat{\beta}_1 = 0.140, \hat{\tau}^2 = 12.069, \hat{\alpha} = 2.050 \) and \( \hat{\gamma} = 2.764 \). For the ‘Health’ data, the estimates of the model parameters are \( \hat{\beta}_0 = 8.819, \hat{\beta}_1 = 0.192, \hat{\tau}^2 = 5.497, \hat{\alpha} = 9.502 \) and \( \hat{\gamma} = 2.109 \). As seen from the tables, the ‘Education’ data have more variability than the ‘Health’ data. The approximated empirical Bayes estimator \( \hat{\xi}_{\text{AEB}}i \) returns almost similar values as the empirical Bayes estimator \( \hat{\xi}_{\text{EB}}i \) does. Both estimators do not shrink \( y_i \) so much. It is also from Table 4 that \( \hat{\xi}_{\text{AEB}}i \) and \( \hat{\xi}_{\text{EB}}i \) shrink \( y_i \) more toward \( \hat{\beta}_0 + z_i \hat{\beta}_1 \) when the values of \( V_i \) are larger. The MSE estimates of \( \hat{\xi}_{\text{AEB}}i \) give large values for large \( V_i \)'s.

Table 4: Predicted values and the MSE estimates for the ‘Education’ data

<table>
<thead>
<tr>
<th>Prefecture</th>
<th>( V_i )</th>
<th>( y_i )</th>
<th>( z_i^T \hat{\beta} )</th>
<th>( \hat{\xi}_{\text{AEB}}i )</th>
<th>( \hat{\xi}_{\text{EB}}i )</th>
<th>( \text{MSE}_{\text{AEB}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saitama</td>
<td>72.622</td>
<td>32.608</td>
<td>19.309</td>
<td>27.805</td>
<td>29.064</td>
<td>5.868</td>
</tr>
</tbody>
</table>

Table 5: Predicted values and the MSE estimates for the ‘Health’ data

<table>
<thead>
<tr>
<th>Prefecture</th>
<th>( V_i )</th>
<th>( y_i )</th>
<th>( z_i^T \hat{\beta} )</th>
<th>( \hat{\xi}_{\text{AEB}}i )</th>
<th>( \hat{\xi}_{\text{EB}}i )</th>
<th>( \text{MSE}_{\text{AEB}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ibaraki</td>
<td>1.160</td>
<td>10.351</td>
<td>10.946</td>
<td>10.369</td>
<td>10.410</td>
<td>0.211</td>
</tr>
<tr>
<td>Tochigi</td>
<td>3.964</td>
<td>11.759</td>
<td>11.080</td>
<td>11.720</td>
<td>11.572</td>
<td>0.369</td>
</tr>
<tr>
<td>Gunma</td>
<td>3.444</td>
<td>8.737</td>
<td>10.307</td>
<td>8.818</td>
<td>9.139</td>
<td>0.349</td>
</tr>
<tr>
<td>Saitama</td>
<td>0.920</td>
<td>11.133</td>
<td>11.316</td>
<td>11.138</td>
<td>11.146</td>
<td>0.194</td>
</tr>
<tr>
<td>Chiba</td>
<td>3.720</td>
<td>12.808</td>
<td>11.150</td>
<td>12.718</td>
<td>12.388</td>
<td>0.358</td>
</tr>
<tr>
<td>Tokyo</td>
<td>1.161</td>
<td>13.803</td>
<td>10.959</td>
<td>13.714</td>
<td>13.477</td>
<td>0.208</td>
</tr>
<tr>
<td>Kanagawa</td>
<td>0.479</td>
<td>14.496</td>
<td>11.088</td>
<td>14.411</td>
<td>14.293</td>
<td>0.163</td>
</tr>
</tbody>
</table>

6 Concluding Remarks

In the Fay-Herriot random dispersion (FHRD) model, we have derived the approximated empirical Bayes (AEB) estimator with the closed form and provided the second-order
unbiased estimator of the MSE of the AEB estimator via the single parametric bootstrap method. Through various simulation experiments and empirical studies, it has been shown that the difference between the AEB estimator and the empirical Bayes estimator given in Maiti, et al. (2014) is small. This means that the AEB estimator is useful irrespective of validity of the approximation. It has been observed that the estimators suggested in this paper for the model parameters have good performances. Especially, our estimator of \( \alpha \) is described in the closed form, and it performs well.

Acknowledgments.

Research of the second author was supported in part by Grant-in-Aid for Scientific Research (23243039 and 26330036) from Japan Society for the Promotion of Science.

A Appendix

We here give proofs of Proposition 2.1 and Theorems 3.1 and 4.1.

[1] Proof of Proposition 2.1. It follows from (2.5) and (2.6) that

\[
E \left[ \frac{1}{\tau^2 \eta_i + 1} \mid y_i, V_i \right] = \frac{\int_0^\infty (\tau^2 \eta_i + 1)^{-1} D_i(\eta_i) g^*_i(\eta_i) d\eta_i}{\int_0^\infty D_i(\eta_i) g^*_i(\eta_i) d\eta_i}, \]

where

\[
g^*_i(\eta_i) = \eta_i^{m_i + 1 + \alpha - 1} \exp[-\eta_i(V_i + \gamma)/2],
\]

\[
D_i(\eta_i) = \frac{1}{\sqrt{\tau^2 \eta_i + 1}} \exp \left[ - \frac{\eta_i}{2(\tau^2 \eta_i + 1)} (y_i - z_i^T \beta)^2 \right].
\]

It is noted that \( D_i(\eta_i) \) is a decreasing function of \( \eta_i \). Then, we first show that

\[
\frac{\int_0^\infty (\tau^2 \eta_i + 1)^{-1} D_i(\eta_i) g^*_i(\eta_i) d\eta_i}{\int_0^\infty D_i(\eta_i) g^*_i(\eta_i) d\eta_i} \geq \frac{\int_0^\infty (\tau^2 \eta_i + 1)^{-1} g^*_i(\eta_i) d\eta_i}{\int_0^\infty g^*_i(\eta_i) d\eta_i}.
\] (A.1)

This inequality is equivalent to

\[
E^*_\{ (\tau^2 \eta_i + 1)^{-1} D_i(\eta_i) \} \geq E^*_\{ (\tau^2 \eta_i + 1)^{-1} \} E^*_\{ D_i(\eta_i) \},
\] (A.2)

where \( E^*_\{ \cdot \} \) is the expectation with respect to the pdf \( C_i g^*_i(\eta_i) \) for some constant \( C_i \). The inequality (A.2) is equivalent to

\[
\text{Cov}^*_\{ (\tau^2 \eta_i + 1)^{-1}, D_i(\eta_i) \} = E^*_\{ \{ (\tau^2 \eta_i + 1)^{-1} - E^*_\{ (\tau^2 \eta_i + 1)^{-1} \} \} \{ D_i(\eta_i) - E^*_\{ D_i(\eta_i) \} \} \} \geq 0,
\]

which holds since both \( (\tau^2 \eta_i + 1)^{-1} \) and \( D_i(\eta_i) \) are decreasing in \( \eta_i \). Thus, one gets the inequality (A.1).
Since $(\tau^2\eta_i + 1)^{-1}$ is a convex function of $\eta_i$, the Jensen inequality is applied to show the inequality
\[
\frac{\int_0^{\infty} (\tau^2\eta_i + 1)^{-1} g_i^* (\eta_i) d\eta_i}{\int_0^{\infty} g_i^*(\eta_i) d\eta_i} \geq \left( \frac{\tau^2}{\int_0^{\infty} g_i^*(\eta_i) d\eta_i} \right)^{-1}.
\]
Noting that $g_i^*(\eta_i)$ is proportional to the pdf of $\text{Ga}((n_i + 1 + \alpha)/2, 2/(V_i + \gamma))$, we can see that $\int_0^{\infty} \eta_i g_i^*(\eta_i) d\eta_i / \int_0^{\infty} g_i^*(\eta_i) d\eta_i = (n_i + \alpha + 1)/(V_i + \gamma)$. Namely,
\[
\frac{\int_0^{\infty} (\tau^2\eta_i + 1)^{-1} g_i^*(\eta_i) d\eta_i}{\int_0^{\infty} g_i^*(\eta_i) d\eta_i} \geq \frac{1}{\tau^2(n_i + \alpha + 1)/(V_i + \gamma) + 1}.
\]
Combining (A.1) and (A.3) proves Proposition 2.1.

[2] Proof of Theorem 3.1. For notational simplicity, let $\mathbf{\theta} = (\theta_1, \theta_2, \theta_3)^T = (\tau^2, \alpha, \gamma)^T$ and $\tilde{\mathbf{\theta}} = (\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3)^T = (\tilde{\tau}^2, \tilde{\alpha}, \tilde{\gamma})^T$. Since $\tilde{\mathbf{\beta}}$ is expressed as $\tilde{\mathbf{\beta}} = \mathbf{\beta}(\mathbf{\theta})$, the Taylor series expansion gives that
\[
\tilde{\mathbf{\beta}} - \mathbf{\beta} = \tilde{\mathbf{\beta}} - \mathbf{\beta} + \sum_{a=1}^{3} \left\{ \left( \sum_{j=1}^{m} (1 - B_j) z_j \right)^{-1} \left( \sum_{j=1}^{m} \frac{\partial B_j}{\partial \theta_a} z_j \right) \right\} (\tilde{\theta}_a - \theta_a) + o_p(m^{-1}),
\]
where $\mathbf{\beta}^* = \{ \sum_{j=1}^{m} (\partial B_j / \partial \theta_a) z_j z_j^T \}^{-1} \sum_{j=1}^{m} (\partial B_j / \partial \theta_a) z_j y_j$. Since $E[(y_i - z_i^T \mathbf{\beta})^2 | V_i] = (V_i + \gamma)/(n_i + \alpha - 2) + \tau^2$ from (2.13), it is observed that
\[
E[(\tilde{\mathbf{\beta}} - \mathbf{\beta})(\tilde{\mathbf{\beta}} - \mathbf{\beta})^T | V_i] = E \left[ \left( \sum_{j=1}^{m} (1 - B_j) z_j \right)^{-1} \left( \sum_{j=1}^{m} z_j z_j^T (1 - B_j)^2 \left\{ \frac{V_j + \gamma}{n_j + \alpha - 2} + \tau^2 \right\} \right) \times \left( \sum_{j=1}^{m} (1 - B_j) z_j \right)^{-1} \right] | V_i,
\]
which implies that $E[(\tilde{\mathbf{\beta}} - \mathbf{\beta})(\tilde{\mathbf{\beta}} - \mathbf{\beta})^T] = O(m^{-1})$, and $\tilde{\mathbf{\beta}} - \mathbf{\beta} = O_p(m^{-1/2})$ under the condition (A) for $n_j + \alpha > 2$. Also, it is seen that
\[
E[(\tilde{\mathbf{\beta}} - \mathbf{\beta})(\tilde{\mathbf{\beta}} - \mathbf{\beta})^T | y_i, V_i] = E[(\tilde{\mathbf{\beta}} - \mathbf{\beta})(\tilde{\mathbf{\beta}} - \mathbf{\beta})^T] + E \left[ \left( \sum_{j=1}^{m} (1 - B_j) z_j z_j^T \right)^{-1} z_j z_j^T (1 - B_j)^2 \left( (y_i - z_i^T \mathbf{\beta})^2 - \left\{ \frac{V_j + \gamma}{n_j + \alpha - 2} + \tau^2 \right\} \right) \times \left( \sum_{j=1}^{m} (1 - B_j) z_j z_j^T \right)^{-1} | y_i, V_i \right],
\]
19
which shows that \( E[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)^T | y, V_i] = E[(\bar{\beta} - \beta)(\bar{\beta} - \beta)^T] + O_p(m^{-2}) \) and 
\( \tilde{\beta} - \beta | y_i, V_i = O_p(m^{-1/2}) \). Clearly, \( E[\beta - \beta] = 0 \), and it is observed that

\[
E[\tilde{\beta} - \beta | y_i, V_i] = E\left[\sum_{j=1}^{m} (1 - B_j)z_jz_j^T \right]^{-1} \sum_{j=1}^{m} (1 - B_j)z_j(y_j - z_j^T \beta) | y_i, V_i
\]

\[
= E\left[\sum_{j=1}^{m} (1 - B_j)z_jz_j^T \right]^{-1} (1 - B_i)z_i(y_i - z_i^T \beta),
\]

which means that \( E[\tilde{\beta} - \beta | y_i, V_i] = O_p(m^{-1}) \). Similarly, we can show that \( E[\beta^* - \beta | y_i, V_i] = O_p(m^{-1}) \) and \( \beta^* - \beta | y_i, V_i = O_p(m^{-1/2}) \). Thus, from (A.4), we can verify that

\[
E[(\beta - \beta)(\beta - \beta)^T] = O(m^{-1}),
\]

\[
E[\beta - \beta | y_i, V_i] = +o_p(m^{-1}) \text{ and } E[(\beta - \beta)(\beta - \beta)^T | y_i, V_i] = E[(\beta - \beta)(\beta - \beta)^T] + o_p(m^{-1}),
\]

provided \( \theta = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)^T \) satisfy the results given in Theorem 3.1.

We now prove the results concerning \( \hat{\theta} \) in Theorem 3.1. Let \( F(\theta) = (F_1(\theta), F_2(\theta), F_3(\theta))^T \), where

\[
F_1(\theta) = \frac{1}{m} \sum_{j=1}^{m} \left\{ \frac{\alpha}{\gamma} \tau^2 + \frac{1}{n_j + \alpha - 2} - \frac{(y_j - z_j^T \hat{\theta}_{OLS})^2}{V_j + \gamma} \right\},
\]

\[
F_2(\theta) = \frac{1}{m} \sum_{j=1}^{m} \left\{ \frac{\alpha^2 V_j}{V_j + \gamma} \log(V_j + \gamma) + \alpha n_j \frac{V_j - \gamma}{V_j + \gamma} \log(V_j + \gamma) - \frac{n_j^2 \gamma}{V_j + \gamma} \log(V_j + \gamma) - 2n_j \right\},
\]

\[
F_3(\theta) = \frac{1}{m} \sum_{j=1}^{m} \left\{ \frac{V_j}{V_j + \gamma} - \frac{n_j}{n_j + \alpha} \right\}.
\]

Since \( F(\hat{\theta}) = 0 \), the consistency of \( \hat{\theta} \) follows from the Cramer method explained in Jiang (2010). It is noted that for \( a = 1, 2, 3 \),

\[
F_a(\theta) = F_a(\theta) + \left( \frac{\partial F_a(\theta)}{\partial \theta^T} \right) (\hat{\theta} - \theta) + \frac{1}{2} (\hat{\theta} - \theta)^T \frac{\partial^2 F_a(\theta)}{\partial \theta \partial \theta^T} (\hat{\theta} - \theta) + o_p(||\hat{\theta} - \theta||^2),
\]

which yields

\[
\hat{\theta} - \theta = - \left( \frac{\partial F(\theta)}{\partial \theta^T} \right)^{-1} F(\theta) - \frac{1}{2} \left( \frac{\partial F(\theta)}{\partial \theta^T} \right)^{-1} \text{Col}_a \left( \hat{\theta} - \theta \right)^T \frac{\partial^2 F_a(\theta)}{\partial \theta \partial \theta^T} (\hat{\theta} - \theta) + o_p(||\hat{\theta} - \theta||^2),
\]

(A.6)

where \( \text{Col}_a(x_a) = (x_1, x_2, x_3)^T \) and \( ||\hat{\theta} - \theta||^2 = (\hat{\theta} - \theta)^T(\hat{\theta} - \theta) \). Hence, it is sufficient to show that \( E[\{F_a(\theta)\}^2] = O(m^{-1}) \), \( E[\{F_a(\theta)\}^2 | y_i, V_i] = O_p(m^{-1}) \), \( E[F_a(\theta)] = O_p(m^{-1}) \) and \( E[F_a(\theta) | y_i, V_i] = O_p(m^{-1}) \) for each \( a \), and \( \partial F(\theta)/\partial \theta^T \) converges to a positive definite matrix.
Concerning $F_1(\theta)$, note that

$$F_1(\theta) = F_1^*(\beta, \theta) + \frac{2}{m} \sum_{j=1}^{m} \frac{y_j - z_j^T \beta}{V_j + \gamma} \cdot z_j^T (\hat{\beta}_{OLS} - \beta) + O_p(m^{-1}),$$

where

$$F_1^*(\beta, \theta) = \frac{1}{m} \sum_{j=1}^{m} \left\{ \frac{\alpha}{n_j + \alpha} \cdot \frac{1}{n_j + \alpha - 2} - \frac{(y_j - z_j^T \beta)^2}{V_j + \gamma} \right\}.$$ 

It can be verified that $E[[F_1(\theta)]^2] = O(m^{-1}), E[[F_1(\theta)]^2 \mid y_i, V_i] = O_p(m^{-1}), E[F_1(\theta)] = O_p(m^{-1})$ and $E[F_1(\theta) \mid y_i, V_i] = O_p(m^{-1})$ if there exists $E[((y_j - z_j^T \beta)^4)/(V_j + \gamma)^2]$. Similarly, the corresponding properties for $F_2(\theta)$ and $F_3(\theta)$ can be demonstrated if there exists $E[[\log(V_j + \gamma)]^2]$. Thus, we need to check these moments. It is noted that

$$E[((y_j - z_j^T \beta)^4 \mid V_j) = E[((y_j - \xi_j)^4 + 6(y_j - \xi_j)^2(\xi_j - z_j^T \beta)^2 + (\xi_j - z_j^T \beta)^4 \mid V_j]$$

$$= E\left[ \frac{3}{n_j} \frac{6\tau^2}{\eta_j} + \frac{3\tau^4}{V_j} \right]$$

$$= \frac{3(V_j + \gamma)^2}{(n_j + \alpha - 2)(n_j + \alpha - 4) + \frac{6\tau^2(V_j + \gamma)}{n_j + \alpha - 2} + 3\tau^4},$$

so that $E[((y_j - z_j^T \beta)^4/(V_j + \gamma)^2]$ is finite if $n_j + \alpha > 4$. To investigate the existence of $E[[\log(V_j + \gamma)]^2]$, we calculate both sides of $4 E[(-\partial^2/\partial \alpha^2) \log f(V_j)] = 4E[((\partial/\partial \alpha) \log f(V_j)]^2]$. The RHS is equal to $E[\{\psi((n_j + \alpha)/2) - \psi(\alpha/2) + \log \gamma - \log(V_j + \gamma)^2\}^2 = \{\psi((n_j + \alpha)/2) - \psi(\alpha/2) + \log \gamma^2 - E[\{\log(V_j + \gamma)^2]\]$. On the other hand, the LHS is $-\psi'(n_j + \alpha/2) + \psi'(\alpha/2)$. Hence,

$$E[[\log(V_j + \gamma)]^2] = \{\psi((n_j + \alpha)/2) - \psi(\alpha/2) + \log \gamma^2 + \psi'((n_j + \alpha)/2) - \psi'(\alpha/2),$$

which is finite for $\alpha > 0$.

Finally, we show that every entry of the matrix $\partial F(\theta)/\partial \theta^T$ converges in probability. For notational simplicity, let $F_{a(i)}(\theta) = \partial F_a(\theta)/\partial \theta_i$ for $a = 1, 2, 3$ and $i = 1, 2, 3$. Then, for $F_{1(i)}$, we have $F_{1(1)} = m^{-1} \sum_j \alpha/\{\gamma(n_j + \alpha)\}$, and

$$F_{1(2)} = \frac{1}{m} \sum_j \left\{ \frac{n_j \tau^2 / \gamma}{(n_j + \alpha)^2} - \frac{1}{(n_j + \alpha - 2)^2} \right\}, \quad F_{1(3)} = \frac{1}{m} \sum_j \left\{ - \frac{\tau^2 / \gamma^2}{n_j + \alpha} + \frac{(y_j - z_j^T \hat{\beta}_{OLS})^2}{(V_j + \gamma)^2} \right\}.$$

For $F_{2(i)}$, we have $F_{2(1)} = 0$, and

$$F_{2(2)} = \frac{1}{m} \sum_j \frac{\log(V_j + \gamma) V_j}{V_j + \gamma} \{2\alpha V_j + n_j(V_j - \gamma)\},$$

$$F_{2(3)} = -\frac{1}{m} \sum_j \frac{V_j \log(V_j + \gamma)}{(V_j + \gamma)^2} (\alpha + n_j)^2 + \frac{1}{m} \sum_j \frac{\alpha + n_j}{(V_j + \gamma)^2} \{\alpha V_j - n_j \gamma\}.$$
For $F_{3(i)}$, we have $F_{3(1)} = 0$, and

$$F_{3(2)} = \frac{1}{m} \sum_{j} \frac{n_j}{(n_j + \alpha)^2}, \quad F_{3(3)} = -\frac{1}{m} \sum_{j} \frac{V_j}{(V_j + \gamma)^2}.$$  

Thus, it can be seen that these converge to the limiting values of their expectations if there exist the moments $E[(y_j - z_i^T \beta)^4/(V_j + \gamma)^2]$ and $E[(\log(V_j + \gamma))^2]$. are finite. Therefore, the proof of Theorem 3.1 is complete.

[3] Proof of Theorem 4.1. It is noted from (2.4) that $\xi_i^B = E[\xi_i \mid y_i, V_i] = z_i^T \beta + (1 - E_i)(y_i - z_i^T \beta)$ for $E_i = E[\gamma \eta_i + 1] \mid y_i, V_i$. Then, $J$ can be rewritten as

$$J = E\left[\left(\hat{\xi}_i^{AB} - \xi_i^B\right)\left(\eta_i - \sum_{j=1} w_j \hat{\xi}_j^{AB}\right)\right]$$

$$= E\left[(B_i - E_i)(y_i - z_i^T \beta) \sum_{j=1} w_j B_j(y_j - z_j^T \beta)\right]$$

$$+ E\left[\left(\hat{B}_i z_i^T (\hat{\beta} - \beta) - (\hat{B}_i - B_i)(y_i - z_i^T \beta)\right) \sum_{j=1} w_j \hat{B}_j(y_j - z_j^T \beta)\right]$$

$$+ E\left[\left(B_i - E_i\right)\left(y_i - z_i^T \beta\right) \sum_{j=1} w_j \left(\hat{B}_j - B_j\right)\left(y_j - z_j^T \hat{\beta}\right) - B_j z_j^T (\hat{\beta} - \beta)\right]\right]$$

$$= J_1 + J_2 + J_3. \quad (say)$$

Noting that $E_i$ is a function of $V_i$ and $(y_i - z_i^T \beta)^2$, we can see that

$$E[(B_i - E_i)(y_i - z_i^T \beta)(y_j - z_j^T \beta) \mid V_i, V_j] = \begin{cases} \frac{(B_i - E_i)(y_i - z_i^T \beta)^2}{2} & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases}$$

Thus, $J_1$ is evaluated as

$$J_1 = E\left[w_i B_i (B_i - E_i)(y_i - z_i^T \beta)^2 + (B_i - E_i) \sum_{j \neq i} w_j B_j(y_i - z_i^T \beta)(y_j - z_j^T \beta)\right]$$

$$= E[w_i B_i (B_i - E_i)(y_i - z_i^T \beta)^2],$$

which is of order $O(1)$. Noting that $\hat{\beta} - \beta = O_p(m^{-1/2})$ and $\hat{B}_i - B_i = O_p(m^{-1/2})$, we can demonstrate that

$$J_2 = E\left[\left(B_i z_i^T (\hat{\beta} - \beta) - (\hat{B}_i - B_i)(y_i - z_i^T \beta)\right) \sum_{j=1}^m w_j B_j(y_j - z_j^T \beta)\right] + o(1)$$

$$J_3 = E\left[(B_i - E_i)(y_i - z_i^T \beta) \sum_{j=1}^m w_j \left(\hat{B}_j - B_j\right)\left(y_j - z_j^T \hat{\beta}\right) - B_j z_j^T (\hat{\beta} - \beta)\right] + o(1).$$

22
It is easily seen that

$$E \left[ \left\{ \sum_{j=1}^{m} w_j B_j (y_j - z_j^T \beta) \right\}^2 \mid V_1, \ldots, V_m \right] = E \left[ \sum_{j=1}^{m} w_j^2 B_j^2 (y_j - z_j^T \beta)^2 \mid V_1, \ldots, V_m \right],$$

which implies that $\sum_{j=1}^{m} w_j B_j (y_j - z_j^T \beta) = O_p(m^{-1/2})$. Thus, one gets that $J_2 = O(1)$.

Concerning $J_3$, it is noted that $\hat{\tau}^2$ given in (2.15) is approximated as

$$\hat{\tau}^2 = \left( \sum_{i=1}^{m} \frac{\alpha / \gamma}{n_i + \alpha} \right)^{-1} \sum_{i=1}^{m} \left\{ \frac{(y_i - z_i^T \beta)^2}{V_i + \gamma} - \frac{1}{n_i + \alpha - 2} \right\} + O_p(m^{-1/2}),$$

so that we can regard $\hat{B}_i$ as a function of $(y_i - z_i^T \beta)^2$ and $V_i, i = 1, \ldots, m$. Hence,

$$E \left[ (B_i - E_i)(y_i - z_i^T \beta) \sum_{j=1}^{m} w_j (\hat{B}_j - B_j)(y_j - z_j^T \beta) \mid V_1, \ldots, V_m \right] = E[ (B_i - E_i)(y_i - z_i^T \beta)^2 w_i(\hat{B}_i - B_i)],$$

which is of $O(m^{-1/2})$. Since $\hat{\beta} - \bar{\beta} = O_p(m^{-1/2})$, it is seen that

$$E \left[ (B_i - E_i)(y_i - z_i^T \beta) \sum_{j=1}^{m} w_j B_j z_j^T (\bar{\beta} - \beta) \mid V_1, \ldots, V_m \right]$$

$$= E \left[ (B_i - E_i)(y_i - z_i^T \beta) \sum_{j=1}^{m} w_j B_j z_j^T \left( \sum_{a=1}^{m} (1 - B_a) z_a z_a^T \right)^{-1} \right. \times \sum_{a=1}^{m} (1 - B_a) z_a (y_a - z_a^T \beta) \mid V_1, \ldots, V_m \right]$$

$$= E \left[ (B_i - E_i)(y_i - z_i^T \beta)^2 \sum_{j=1}^{m} w_j B_j z_j^T \left( \sum_{a=1}^{m} (1 - B_a) z_a z_a^T \right)^{-1} (1 - B_i) z_i \mid V_1, \ldots, V_m \right],$$

which is of order $O_p(1)$. Hence, it is concluded that $J_3 = O(1)$. Therefore, the proof of Theorem 4.1 is complete.

References


