Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://amstat.tandfonline.com/loi/uasa20

Multi-Goal Prior Selection: A Way to Reconcile Bayesian and Classical Approaches for Random Effects Models

Masayo Y. Hirose & Partha Lahiri

To cite this article: Masayo Y. Hirose & Partha Lahiri (2020): Multi-Goal Prior Selection: A Way to Reconcile Bayesian and Classical Approaches for Random Effects Models, Journal of the American Statistical Association, DOI: 10.1080/01621459.2020.1737532

To link to this article: https://doi.org/10.1080/01621459.2020.1737532

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2020, VOL. 00, NO. 0, 1–11: Theory and Methods
https://doi.org/10.1080/01621459.2020.1737532

Multi-Goal Prior Selection: A Way to Reconcile Bayesian and Classical Approaches for Random Effects Models

Masayo Y. Hirose[a] and Partha Lahiri[b]

[a]The Institute of Mathematics for Industry, Kyushu University, Fukuoka, Japan; [b]The Joint Program in Survey Methodology & Department of Mathematics, University of Maryland, College Park, MD

ABSTRACT

The two-level normal hierarchical model has played an important role in statistical theory and applications. In this article, we first introduce a general adjusted maximum likelihood method for estimating the unknown variance component of the model and the associated empirical best linear unbiased predictor of the random effects. We then discuss a new idea for selecting prior for the hyperparameters. The prior, called a multi-goal prior, produces Bayesian solutions for hyperparmeters and random effects that match (in the higher order asymptotic sense) the corresponding classical solution in linear mixed model with respect to several properties. Moreover, we establish for the first time an analytical equivalence of the posterior variances under the proposed multi-goal prior and the corresponding parametric bootstrap second-order mean squared error estimates in the context of a random effects model.

ARTICLE HISTORY
Received February 2019
Accepted February 2020

KEYWORDS
Adjusted maximum likelihood method; Empirical Bayes; Empirical best linear unbiased prediction; Linear mixed model.

1. Introduction

Simultaneous estimation of several independent normal means has been a topic of great research interest, especially in the 60s, 70s, and 80s, after the publication of the celebrated James–Stein estimator (James and Stein 1961). Let $y = (y_1, \ldots, y_m)'$ be a maximum likelihood estimator of $\theta = (\theta_1, \ldots, \theta_m)'$ under the model: $y_i|\theta_i \overset{\text{ind.}}{\sim} N(\theta_i, 1)$, $i = 1, \ldots, m$. James–Stein (1961) provided a surprising result that for $m \geq 3$, y is an inadmissible estimator of θ under the model and the sum of squared error loss function: $L(\hat{\theta}, \theta) = \sum_{i=1}^{m}(\hat{\theta}_i - \theta_i)^2$. They also showed that the estimator $\hat{\theta}_i^{\text{JS}} = (1 - \hat{B}^{\text{JS}})y_i$, where $\hat{B}^{\text{JS}} = (m-2)/(\sum_{i=1}^{m} y_i^2)$, dominates y in terms of the frequentist's risk. To be specific, $E[\sum_i^m(\hat{\theta}_i^{\text{JS}} - \theta_i)^2|\theta] \leq E[\sum_i^m(y_i - \theta_i)^2|\theta]$, for all $\theta \in \mathcal{R}^m$, the m-dimensional Euclidean space, with strict inequality holding for at least one point θ.

The potential of different extensions of the James–Stein estimator to improve data analysis became transparent when Efron and Morris (1973) provided an empirical Bayesian justification of the James–Stein estimator using the prior $\theta_i \overset{\text{iid}}{\sim} N(0, A)$, $i = 1, \ldots, m$. Some earlier applications of empirical Bayesian method include the estimation of (a) false alarm probabilities in New York City (Carter and Rolph 1974), (b) batting averages of major league baseball players (Efron and Morris 1975), (c) prevalence of toxoplasmosis in El Salvador (Efron and Morris 1975), and (d) per-capita income of small places in the USA (Fay and Herriott 1979). More recently, variants of the method given in Efron and Morris (1973) was used to estimate poverty rates for the US states, counties, and school districts (Citro and Kalton 2000) and Chilean municipalities (Casas-Cordero, Encina, and Lahiri 2016), and to estimate proportions at the lowest level of literacy for states and counties (Mohadjer et al. 2012).

The following two-level Normal hierarchical model is an extension of the model used by Efron and Morris (1973):
For $i = 1, \ldots, m$,

Level 1. (sampling model): $y_i|\theta_i \overset{\text{ind.}}{\sim} N(\theta_i, D_i)$;

Level 2. (linking model): $\theta_i \overset{\text{ind.}}{\sim} N(x_i'\beta, A)$.

In the above model, level 1 is used to account for the sampling distribution of unbiased estimates y_i based on observations taken from the ith population. In this model, we assume that the sampling variances D_i are known and this assumption often follows from the asymptotic variances of transformed direct estimates (Carter and Rolph 1974; Efron and Morris 1975) or from empirical variance modeling (Fay and Herriot 1979; Otto and Bell 1995). Level 2 links the random effects θ_i to a vector of p known auxiliary variables $x_i = (x_{i1}, \ldots, x_{ip})'$, which are often obtained from various alternative data sources. The parameters β and A are generally unknown and are estimated from the available data. We assume that $\beta \in \mathcal{R}^p$, the p-dimensional Euclidian space. In the growing field of small area estimation, this model is commonly referred to as the Fay–Herriot model, named after the authors of the landmark paper with more than 1290 citations to date (according to Google Scholar) by Fay and Herriot (1979). For a comprehensive review of small area estimation, the readers are referred to Lahiri (2003), Jiang (2007), and Rao and Molina (2015).

We may be interested in the high-dimensional parameters (random effects) θ_i and/or the hyperparameters β and A. The

CONTACT Masayo Y. Hirose masayo@imi.kyushu-u.ac.jp The Institute of Mathematics for Industry, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, Japan 8190395.
A part of the research was conducted when the first author was an assistant professor at the Institute of Statistical Mathematics, Tokyo.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/JASA.

estimation problem can be addressed using either Bayesian or linear mixed model classical approach. When hyperparameters are known, both the Bayesian and linear mixed model classical approaches use conditional distribution of θ_i given the data for point estimation and measuring uncertainty of the point estimator. To elaborate, the posterior mean of θ_i, the Bayesian point estimator, is identical to the best predictor of θ_i. Moreover, the posterior variance of θ_i is identical to the mean squared error of the best predictor. When A is known but β is unknown, a flat prior is generally assumed for β under the Bayesian approach. Interestingly, in this unknown β case, the posterior mean and posterior variance of β are identical to the maximum likelihood estimator of β and the variance of the maximum likelihood estimator, respectively. Moreover, the posterior mean and variance of θ_i are identical to the best linear unbiased predictor of θ_i and its mean squared error, respectively.

When both β and A are unknown, flat prior, that is, $\pi(\beta, A) \propto 1$, $\beta \in \mathcal{R}^p, A > 0$, is common though a few other priors for A have been considered (see, e.g., Datta, Rao, and Smith 2005; Morris and Tang 2011). In a linear mixed model classical approach, different estimators of A have been proposed and the estimator of β is obtained by plugging in an estimator of A in the maximum likelihood estimator of β when A is known. In this general case, the relationship between the Bayesian and linear mixed model classical approach is not clear. The main goal of this article is to understand the nature of such relationship. In particular, we answer the following question: For a given classical method of estimation of A, is it possible to find a prior on A that will make the Bayesian solution closer to the classical solution in achieving desirable multiple goals? In this article, we set the desirable multiple goals as (i)–(v), given below, in terms of probability, up to the order of $O_p(m^{-1})$.

(i) The posterior mean of the shrinkage parameter $B_i = D_i/(A + D_i)$ is identical to its desirable classical estimator.
(ii) The posterior variance of the shrinkage parameter is identical to the variance of its classical estimator.
(iii) The posterior mean of the random effect θ_i is identical to the empirical best linear unbiased predictor given in Hirose and Lahiri (2018).
(iv) The posterior variance of the random effect is identical to the Taylor series second-order mean squared error estimators given in Hirose and Lahiri (2018).
(v) The posterior variance of the random effect is identical to the parametric bootstrap second-order mean squared error estimators proposed by Hirose and Lahiri (2018).

These desirable multiple goals are described in details in Section 3. A subset of these goals is given in Theorem 2.

Let us now explain these multiple goals (i)–(v). To this end, we first note that Morris and Tang (2011) pointed out the need for accurately estimating the shrinkage parameters $B_i = D_i/(A + D_i)$ as they appear linearly in the Bayes estimators of θ_i, which are the prime parameters of interest in many applications like the small area estimation. Moreover, the shrinkage parameters are good indicators of the strength of the prior on the random effects θ_i. Despite the importance of shrinkage parameters, relatively little research has been conducted to understand the theoretical properties of existing estimators. For the balanced case when $D_i = D$, $i = 1, \ldots, m$, Morris (1983) proposed an exact unbiased estimator of $B = D/(A + D)$ and showed component-wise dominance of the resulting empirical Bayes estimator of θ_i under the joint distribution of $\{(y_i, \theta_i),\ i = 1, \ldots, m\}$ when $p \leq m - 3$. For the general unbalanced case, Hirose and Lahiri (2018) proposed an adjusted maximum likelihood estimator of B_i that satisfies the following desirable properties: First, the method yields an estimator of B_i that is strictly less than 1, which prevents the overshrinking problem in the related empirical best linear unbiased predictor of θ_i. Second, this adjusted maximum likelihood estimator of B_i has the smallest bias among all existing rival estimators in the higher order asymptotic sense. Third, when this adjusted maximum likelihood method is used, second-order unbiased estimator of mean squared error of empirical best linear unbiased predictor can be produced in a straightforward way without additional bias corrections that are necessary for other existing variance component estimation methods. For prior work on the adjusted maximum likelihood method, the readers are referred to Lahiri and Li (2009), Li and Lahiri (2010), Yoshimori and Lahiri (2014a, 2014b), Hirose and Lahiri (2018), and Hirose (2017, 2019).

As stated in Morris and Tang (2011), flat prior leads to admissible minimax estimators of the random effects for a special case of the model. In Section 3, we show that the bias of the Bayes estimator of B_i, under the flat prior and the two-level model, is $O(m^{-1})$ except for the balanced case when it is of lower order $o(m^{-1})$. Thus, in general, the Bayes estimator of B_i, under the flat prior, has more bias than the adjusted maximum likelihood estimator of Hirose and Lahiri (2018) in the higher order asymptotic sense. In this section, we propose a prior for the hyperparameters that leads to the Bayes estimator of B_i with bias of lower order $o(m^{-1})$ and thus is on par with the adjusted maximum likelihood of Hirose and Lahiri (2018). Interestingly, this prior also makes the resulting Bayesian method much closer to the Hirose–Lahiri's empirical best linear unbiased prediction method in multiple sense. In particular, the posterior variance of the random effect θ_i, under the proposed prior, is identical to both the Taylor series and parametric bootstrap second-order mean squared error estimators of Hirose and Lahiri (2018) in the higher order asymptotic sense. To our knowledge, we establish for the first time the relationship between the Bayesian posterior variance and parametric bootstrap mean squared error estimator in this higher-order asymptotic sense.

The outline of the article is as follows. In Section 2, we first introduce a classical method for the two level model by proposing a general adjustment factor in estimating A. We show how the method is related to the commonly used residual maximum likelihood method for a given choice of the adjustment factor. We then construct a prior, called a multi-goal prior, that provides a Bayesian solution close (with respect to several properties in higher order asymptotic sense) to classical solution to estimate the hyperparameters and random effects. Section 3 discusses prior choice for an important special case considered by Hirose and Lahiri (2018). In addition to the multiple properties discussed in Section 2, this section develops a unique multi-goal prior that establishes a relationship of the posterior variances of the random effects with the Hirose–Lahiri Taylor series and parametric bootstrap mean squared error estimators

that do not require the usual complex bias corrections. We reiterate that this article demonstrates for the first time how to bring the Bayesian and classical parametric bootstrap methods closer in the context of random effects models. In Section 4, we compare the proposed multi-goal prior with the superharmonic prior using a real life data. In Section 5, we discuss issues in extending our results to a general model. All the technical proofs are deferred to the Appendix.

2. Prior Choice for Reconciliation of the Bayesian and Classical Approach

In this section, we first introduce a general classical method for estimation of hyperparameters and random effects in the two-level Normal hierarchical model. Then we construct prior for the hyperparameters so that the corresponding Bayesian method is identical to the classical method in the higher order asymptotic sense with respect to multiple properties.

We first introduce the empirical best linear unbiased predictor of θ_i when the variance component A is estimated by a general adjusted maximum likelihood method. To this end, we define mean squared error of a given predictor $\hat{\theta}_i$ of θ_i as $M_i(\hat{\theta}_i) = E(\hat{\theta}_i - \theta_i)^2$, where the expectation is with respect to the joint distribution of $y = (y_1, \ldots, y_m)'$ and $\theta = (\theta_1, \ldots, \theta_m)'$ under the two-level normal model. The best linear unbiased predictor $\hat{\theta}_i^{\text{BLUP}}$ of θ_i, which minimizes $M_i(\hat{\theta}_i)$ among all linear unbiased predictors $\hat{\theta}_i$, is given by $\hat{\theta}_i^{\text{BLUP}}(A) = (1 - B_i)y_i + B_i x_i' \hat{\beta}(A)$, where $B_i \equiv B_i(A) = D_i/(A + D_i)$ is the shrinkage factor and $\hat{\beta}(A) = (X'V^{-1}X)^{-1}X'V^{-1}y$ is the weighted least square estimator of β when A is known. In this formula, $X' = (x_1, \ldots, x_m)$ denotes $p \times m$ matrix of known auxiliary variables and $V = \text{diag}(A + D_1, \ldots, A + D_m)$ denotes a $m \times m$ diagonal covariance matrix of y.

We consider the following general adjusted maximum likelihood estimator $\hat{A}_{i;G}$ of A:

$$\hat{A}_{i;G} = \arg\max_{A \geq 0} h_{i;G}(A) L_{\text{RE}}(A), \tag{1}$$

where $L_{\text{RE}}(A) = |X'V^{-1}X|^{-1/2}|V|^{-1/2}\exp(-y'Py/2)$ is the residual maximum likelihood of A with $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$ and the general adjustment factor $h_{i;G}(A)$ satisfies condition R5 in Appendix A.

Plugging in $\hat{A}_{i;G}$ for A in the best linear unbiased predictor, one obtains an empirical best linear unbiased predictor $\hat{\theta}_i^{\text{EB}}(\hat{A}_{i;G})$ of θ_i. Note that maximum likelihood, residual maximum likelihood and one of the adjusted maximum likelihood estimator proposed in Li and Lahiri (2010) of A can be produced when $h_{i;G}(A) = C$, $C|X'V^{-1}X|^{1/2}$ and CA, respectively, where C is a generic constant free from A. Hereafter, let $\hat{A}_{\text{RE}}$ define as the residual maximum likelihood estimator of A, which can be obtained as

$$\hat{A}_{\text{RE}} = \arg\max_{A \geq 0} L_{\text{RE}}(A). \tag{2}$$

Since the residual maximum likelihood estimator $\hat{A}_{\text{RE}}$ is one of the estimators achieving unbiasedness, up to the order of $O(m^{-1})$ (Das, Jiang, and Rao 2004; Yoshimori and Lahiri 2014a), it is of interest to establish a relationship between the general adjusted maximum likelihood estimator and the residual maximum likelihood estimator. We describe such relationship in Theorem 1; see Appendix A.1 for a proof.

Theorem 1. Under regularity conditions R1–R5 given in Appendix A,

$$\hat{A}_{i;G} - \hat{A}_{\text{RE}} = \frac{2\tilde{l}_{i;G}^{(1)}(A)}{tr[V^{-2}]} + o_p(m^{-1}),$$

where $\tilde{l}_{i;G}^{(1)}(A) = \frac{\partial \log h_{i;G}(A)}{\partial A}$ and $\hat{A}_{i;G}$ and $\hat{A}_{\text{RE}}$ are defined in (1) and (2), respectively.

We now present Theorem 2 for constructing a prior, starting from a given adjustment factor $h_{i;G}(A)$, to bring the resulting Bayesian method closer to the classical method with respect to three criteria. To this end, let $p(\beta, A)$ denote the prior for (β, A). Following Datta, Rao, and Smith (2005), we assume $p(\beta, A) \propto \pi(A)$ and introduce the following notations to be used throughout the article:

$$\hat{b}_1 = \frac{\partial B_i}{\partial A}\Big|_{\hat{A}_{\text{RE}}}, \ \hat{b}_2 = \frac{\partial^2 B_i}{\partial A^2}\Big|_{\hat{A}_{\text{RE}}}, \ \hat{\rho}_1 = \frac{\partial \log \pi(A)}{\partial A}\Big|_{\hat{A}_{\text{RE}}},$$
$$\hat{h}_2 = -\frac{1}{m}\frac{\partial^2 l_{\text{RE}}}{\partial A^2}\Big|_{\hat{A}_{\text{RE}}} = \frac{tr[V^{-2}]}{2m} + o_p(m^{-1}),$$
$$\hat{h}_3 = -\frac{1}{m}\frac{\partial^3 l_{\text{RE}}}{\partial A^3}\Big|_{\hat{A}_{\text{RE}}} = -\frac{2tr[V^{-3}]}{m} + o_p(m^{-1}),$$

where l_{RE} is the logarithm of residual likelihood.

Theorem 2. Under regularity conditions R1–R5, if $p(\beta, A) \propto \pi_{i;G}(A)$ with

$$\pi_{i;G}(A) \propto (A + D_i)tr(V^{-2})h_{i;G}(A), \tag{3}$$

we have

(i) $\hat{B}_i^{\text{GHB}} \equiv E[B_i|y] = \hat{B}_i(\hat{A}_{i;G}) + o_p(m^{-1})$;
(ii) $V[B_i|y] = \text{var}(\hat{B}_i(\hat{A}_{i;G})) + o_p(m^{-1})$;
(iii) $\hat{\theta}_i^{\text{GHB}} \equiv E[\theta_i|y] = \hat{\theta}_i(\hat{A}_{i;G}) + o_p(m^{-1})$,

where GHB stands for general hierarchical Bayes estimator. We call the estimator GHB because it matches general adjusted maximum likelihood method.

The proof of Theorem 2 is deferred to Appendix A.2.

Remark 1. We have several remarks on the prior $\pi_{i;G}(A)$ given by (3).

(a) Theorem 2 is valid for multiple choices of $h_{i;G}$.
(b) There exists at least one strictly positive estimate of A if $h_{i;G}(A) > 0$ and

$$h_{i;G}(A) = o(A^{(m-p)/2}), \tag{4}$$

for large A under R6 and R7.
(c) Note that $h_{i;G}(A)$ may not qualify as a bonafide prior since it may result in an improper posterior (see, e.g., Yoshimori and Lahiri 2014b). However, if we restrict the class of priors to $h_{i;G}(A) = (A + D_i)^s$ for some $s > 0$, we show in Appendix B.1 that $h_{i;G}(A) = o(A^{(m-p-2)/2})$ is a sufficient condition

for the propriety of posterior and hence can serve as a prior for A.

On the other hand, it is straightforward to show that $\pi_{i;G}(A)$ given by (3) with $h_{i;G}(A) = o(A^{(m-p)/2})$ yields proper posterior because of multiplication of $h_{i;G}(A)$ by $(A + D_i)tr(V^{-2})$. In either case, Theorem 2 can facilitate users for selecting an adjustment factor in the empirical best linear unbiased prediction approach or prior in the Bayesian approach.

(d) For all i, each estimator $\hat{A}_{i;G}$, like the residual maximum likelihood estimator $\hat{A}_{\mathrm{RE}}$, is asymptotically normal and asymptotically efficient under the regularity conditions. Moreover, the covariance matrix of $(\hat{A}_{1;G}, \ldots, \hat{A}_{m;G})$ converges to a singular matrix for large m. The proofs are shown in Appendix B.2.

3. Multi-Goal Prior for an Important Special Case

Hirose and Lahiri (2018) put forward a classical approach for an important choice of $h_{i;G}(A)$ that satisfies the following desirable properties under regularity conditions R1–R7:

[1] It is desirable to have a second-order unbiased estimator of B_i, that is, $E(\hat{B}_i) = B_i + o(m^{-1})$.

[2] $0 < \inf_{m\geq 1}\hat{B}_i \leq \sup_{m\geq 1}\hat{B}_i < 1$ (a.s.) for protecting the empirical best linear unbiased predictor from over-shrinking to the regression estimator.

[3] It is desirable to obtain a simple second-order unbiased Taylor series mean squared error estimator of the empirical best linear unbiased predictor without any bias correction; that is, $E[\hat{M}_i(\hat{A}_i)] = M_i(\hat{\theta}_i^{\mathrm{EB}}) + o(m^{-1})$.

[4] It is desirable to produce a strictly positive second-order unbiased single parametric bootstrap mean squared error estimator without any bias-correction,

where $\hat{M}_i(\hat{A}_i)$ denotes an estimator of mean squared error of $\hat{\theta}_i^{\mathrm{EB}}(\hat{A})$.

Let $\hat{A}_{i;\mathrm{MG}}, \hat{B}_{i;\mathrm{MG}}, \hat{\theta}_{i;\mathrm{MG}}^{\mathrm{EB}}, \hat{M}_{i;\mathrm{MG}}, \hat{M}_{i;\mathrm{MG}}^{\mathrm{boot}}$ be the Hirose–Lahiri's estimators of A, B_i, the empirical best linear unbiased predictor of θ_i, Taylor series and parametric bootstrap estimators of the mean squared error of the empirical best linear unbiased predictor, respectively. They are given by

$$\begin{aligned}
&\hat{A}_{i;\mathrm{MG}} = \arg\max_{A>0} \tilde{h}_i(A) L_{\mathrm{RE}}(A),\\
&\hat{B}_{i;\mathrm{MG}} = \hat{B}_i(\hat{A}_{i;\mathrm{MG}}),\ \hat{\theta}_{i;\mathrm{MG}}^{\mathrm{EB}} = \hat{\theta}_i^{\mathrm{EB}}(\hat{A}_{i;\mathrm{MG}}),\\
&\hat{M}_{i;\mathrm{MG}} = \hat{M}_i(\hat{A}_{i;\mathrm{MG}}),\ \hat{M}_{i;\mathrm{MG}}^{\mathrm{boot}} = E_*[\hat{\theta}_i(\hat{A}_{i;\mathrm{MG}}^*, y^*) - \theta_i^*]^2,
\end{aligned}$$

where $\tilde{h}_i(A) = h_+(A)(A + D_i)$ with $m > p + 2$; $h_+(A)$ satisfies conditions R6 and R7 in Appendix A; $\theta_i^* = x_i'\hat{\beta}(\hat{A}_{1;\mathrm{MG}}, \ldots, \hat{A}_{m;\mathrm{MG}}) + u_i^*$ with $u_i^* \sim^{\mathrm{ind.}} N(0, \hat{A}_{i;\mathrm{MG}})$; E_* is the expectation with respect to the two-level Normal hierarchical model with β and A replaced by $\hat{\beta}(\hat{A}_{1;\mathrm{MG}}, \ldots, \hat{A}_{m;\mathrm{MG}})$ and $\hat{A}_{i;\mathrm{MG}}$, respectively. Note that the choice of $h_+(A)$ is not unique in general. One can use the choice given in Yoshimori and Lahiri (2014a).

The following corollary follows from Theorem 1, Hirose and Lahiri (2018) and the fact that $\frac{\partial\hat{\beta}(A)}{\partial A} = O_p(m^{-1/2})$.

Corollary 1. Using the regularity conditions,

(i) $\hat{A}_{i;\mathrm{MG}} - \hat{A}_{\mathrm{RE}} = O_p(m^{-1})$;

(ii) $x_i'\hat{\beta}(\hat{A}_{1;\mathrm{MG}}, \ldots, \hat{A}_{m;\mathrm{MG}}) - x_i'\hat{\beta}(\hat{A}_{\mathrm{RE}}) = o_p(m^{-1})$.

In this section, we suggest a Bayesian approach that is close to the classical approach to achieve multiple goals in the higher-order asymptotic sense. To this end, we seek a multi-goal prior on the hyperparameters (β, A) that satisfies all the following properties simultaneously:

(i) $\hat{B}_i^{\mathrm{HB}} \equiv E[B_i|y] = \hat{B}_{i;\mathrm{MG}} + o_p(m^{-1})$;
(ii) $V[B_i|y] = \mathrm{var}(\hat{B}_{i;\mathrm{MG}}) + o_p(m^{-1})$;
(iii) $\hat{\theta}_i^{\mathrm{HB}} \equiv E[\theta_i|y] = \hat{\theta}_{i;\mathrm{MG}} + o_p(m^{-1})$;
(iv) $V[\theta_i|y] = \hat{M}_{i;\mathrm{MG}} + o_p(m^{-1})$;
(v) $V[\theta_i|y] = \hat{M}_{i;\mathrm{MG}}^{\mathrm{boot}} + o_p(m^{-1})$,

where HB stands for Hierarchical Bayes estimator.

First we prepare the following result, which follows from Corollary 1(i) and Hirose and Lahiri (2018):

$$\begin{aligned}
\hat{B}_i(\hat{A}_{i;\mathrm{MG}}) - \hat{B}_i(\hat{A}_{\mathrm{RE}}) &= (\hat{A}_{i;\mathrm{MG}} - \hat{A}_{\mathrm{RE}})b_1 + o_p(m^{-1})\\
&= \{E[\hat{A}_{i;\mathrm{MG}} - A] - E[\hat{A}_{\mathrm{RE}} - A]\}b_1\\
&\quad + o_p(m^{-1})\\
&= -\frac{2D_i}{tr[V^{-2}](A + D_i)^3} + o_p(m^{-1}), \quad (5)
\end{aligned}$$

where $b_1 = \frac{\partial B_i}{\partial A}$.

If we use the flat prior $\pi(A) \propto 1$, we get the following result using equation (21) of Datta, Rao, and Smith (2005) with $b(A) = B_i(A)$ and Equation (5):

$$\begin{aligned}
E[B_i|y] = \hat{B}_i(\hat{A}_{\mathrm{MG}}) &+ \frac{4D_i}{tr[V^2](A + D_i)^2}\left[\frac{1}{A + D_i} - \frac{tr[V^{-3}]}{tr[V^{-2}]}\right]\\
&+ o_p(m^{-1}).
\end{aligned}$$

This result emphasizes that the flat prior $\pi(A) \propto 1$ cannot achieve Property (i) except for the balanced case ($D_i = D$ for all i). We, therefore, seek a prior $\pi(A)$ to satisfy Property (i), even in unbalanced case. To this end, we also use the following result (6) given in (21) of Datta, Rao, and Smith (2005) with $b(A) = B_i(A)$:

$$E[B_i|y] = \hat{B}_i(\hat{A}_{\mathrm{RE}}) + \frac{1}{2m\hat{h}_2}\left(\hat{b}_2 - \frac{\hat{h}_3}{\hat{h}_2}\hat{b}_1\right) + \frac{\hat{b}_1}{m\hat{h}_2}\hat{\rho}_1 + o_p(m^{-1}). \quad (6)$$

It is evident from Equations (5) and (6) that our desired prior must satisfy the following differential equation, up to the order of $O(m^{-1})$:

$$\frac{1}{2mh_2}\left(b_2 - \frac{h_3}{h_2}b_1\right) + \frac{b_1}{mh_2}\rho_1 = -\frac{2D_i}{\mathrm{tr}[V^{-2}](A + D_i)^3}. \quad (7)$$

Note that the differential equation (7) is equivalent to the following differential equation, up to the order of $O_p(m^{-1})$:

$$\begin{aligned}
\rho_1 = \frac{\partial \log\pi(A)}{\partial A} &= -\frac{mh_2}{b_1}\frac{2D}{\mathrm{tr}[V^{-2}](A + D_i)^3} - \frac{1}{2}\left[\frac{b_2}{b_1} - \frac{h_3}{h_2}\right]\\
&= \frac{2}{A + D_i} - \frac{2\mathrm{tr}[V^{-3}]}{\mathrm{tr}[V^{-2}]}. \quad (8)
\end{aligned}$$

Hence, we obtain a solution to differential equation (8) as follows:

$$\pi(A) \propto (A + D_i)^2 \mathrm{tr}[V^{-2}]. \tag{9}$$

Note that the prior (9) depends on i. To elaborate, the proposed prior distributions for two distinct areas i and j will be different unless the sampling variances D_i and D_j are identical. Therefore, we redefine it as

$$\pi_i(A) \propto (A + D_i)^2 \mathrm{tr}[V^{-2}]. \tag{10}$$

Remark 2. We have several important remarks on the prior (10).

(a) The prior satisfies the rest of Properties (ii)–(v) simultaneously, as shown in Appendix B.3. It is remarkable that $\pi_i(A)$ given by (10) is the unique prior to achieve Properties (i)–(v) simultaneously, up to the order of $O_p(m^{-1})$, since $E[g_{1i}(A)|y] = g_{1i}(\hat{A}_{i;\mathrm{MG}}) + o_p(m^{-1})$ shown in (B.6).

(b) The prior given by Equation (10) reduces to the Stein's super-harmonic prior for the balanced case $D_i = D$, $i = 1, \ldots, m$, up to the order of $O_p(m^{-1})$.

(c) Datta, Rao, and Smith (2005) found the same prior by matching (in a higher order asymptotic sense) expected value of the posterior variance of θ_i with the mean squared error of the empirical best linear unbiased predictor with the residual maximum likelihood estimator used for the variance component A. It is interesting to note that the same prior achieves multiple goals, a fact gone unnoticed.

(d) From the result of Ganesh and Lahiri (2008), the prior

$$\pi(A) \propto \frac{\sum\{1/(A + D_i)^2\}}{\sum \omega_i \{D_i^2/(A + D_i)^2\}}$$

also satisfies $\sum_i^m \omega_i E[V(\theta_i|y) - \mathrm{MSE}[\hat{\theta}_i(\hat{A}_{i;\mathrm{MG}})]] = o(m^{-1})$.

(e) For all i, each estimator $\hat{A}_{i;\mathrm{MG}}$, like the residual maximum likelihood $\hat{A}_{\mathrm{RE}}$, is asymptotically normal and asymptotically efficient. Moreover, the covariance matrix of $(\hat{A}_{1;\mathrm{MG}}, \ldots, \hat{A}_{m;\mathrm{MG}})$ converges to a singular matrix for large m. It follows from Remark 1(d) and the result that $\hat{A}_{i;\mathrm{MG}}$ satisfies the conditions for being $\hat{A}_{i;G}$ under the regularity conditions.

(f) Theorem 2 ensures that the following adjustment factor leads the asymptotic equivalent to the results using the Stein's superharmonic prior, up to the order $O_p(m^{-1})$.

$$h_{i;\mathrm{SH}}(A) = \frac{1}{(A + D_i)\mathrm{tr}(V^{-2})}.$$

We also show that the superharmonic prior does not generally attain the multi-goals because

$$\hat{B}_{i;\mathrm{MG}} - \hat{B}_{i;\mathrm{SH}} = -\frac{4B_i^2}{D_i^2 \mathrm{tr}[V^{-2}]} \left\{ B_i - \frac{D_i \mathrm{tr}[V^{-3}]}{\mathrm{tr}[V^{-2}]} \right\} + o_p(m^{-1}), \tag{11}$$

$$\begin{aligned} \hat{\theta}_{i;\mathrm{MG}} - \hat{\theta}_{i;\mathrm{SH}} &= -(\hat{B}_{i;\mathrm{MG}} - \hat{B}_{i;\mathrm{SH}})(x_i'\hat{\beta} - x_i'\beta) + o_p(m^{-1}), \\ &= \frac{4B_i^2}{D_i^2 \mathrm{tr}[V^{-2}]} \left\{ B_i - \frac{D_i \mathrm{tr}[V^{-3}]}{\mathrm{tr}[V^{-2}]} \right\} (y - x_i'\beta) + o_p(m^{-1}), \end{aligned} \tag{12}$$

where $\hat{B}_{i;\mathrm{SH}}$ and $\hat{\theta}_{i;\mathrm{SH}}$ denote the estimator of shrinkage factor and the empirical best linear unbiased predictor using the adjustment factor $h_{i;\mathrm{SH}}(A)$, respectively. The result (11) and (12) follow from (A.4) given in Appendix A.2 and the fact that $\frac{\partial \hat{\beta}(A)}{\partial A} = O_p(m^{-1/2})$.

4. Data Analysis

In this section, using the 1993 Small Area Income and Poverty Estimates (SAIPE) dataset, we demonstrate that our proposed multi-goal prior (MGP) performs better than the superharmonic prior (SHP) in producing Bayesian solutions closer to the multi-goal classical solutions of Hirose and Lahiri (2018). The SAIPE data we use here is from Bell and Franco (2017), available at *https://www.census.gov/srd/csrmreports/byyear.html*. The data contain direct poverty rates, (y_i), associated sampling variances (D_i), and the four auxiliary variables (x_i) derived from administrative and census data for the 50 states and the District of Columbia. Much has been written about SAIPE over the years. See, for instance, the recent book chapter by Bell, Basel, and Maples (2016). Hirose and Lahiri (2018) and Erciulescu, Franco, and Lahiri (2020) considered the same four auxiliary variables in their data analysis.

First we consider the estimation of the shrinkage parameters B_i for all the states. Figure 1 displays classical multi-goal estimates $\hat{B}_{i;\mathrm{MG}}$ (MGF) and Bayes estimates of B_i under the superharmonic (SHP) and the multi-goal priors (MGP) for all the states arranged in decreasing order of $\hat{B}_{i;\mathrm{MG}}$ with all four auxiliary variables and a dummy variable for the intercept. Note that the Bayes estimate of B_i is an one-dimensional integral, which is approximated by numerical integration using the R function "adaptIntegrate." Overall, the Bayes estimates under the multi-goal prior are closer to the classical estimates (MGF) than the superharmonic prior.

Let $\hat{\theta}_i$ denote the Bayes estimate of the random effect θ_i under the superharmonic (SHP) or the multi-goal prior (MGP) for state i. Figure 2 displays the relative difference of the Bayes estimates $\hat{\theta}_i$ from the corresponding classical multi-goal estimates $\hat{\theta}_{i;\mathrm{MG}}$ defined as

$$\mathrm{RD}(\hat{\theta}_i) = \frac{|\hat{\theta}_i - \hat{\theta}_{i;\mathrm{MG}}|}{\hat{\theta}_{i;\mathrm{MG}}} \times 100$$

for all states i. This figure demonstrates that classical multi-goal estimates are closer to the corresponding Bayes estimates under the multi-goal prior (MGP) than the corresponding Bayes estimates under superharmonic prior (SHP).

Figure 3 overall displays Taylor series (MGF) and parametric bootstrap (PB.MG) mean squared error estimates of Hirose and Lahiri (2018) and the posterior variances under the two different priors—MGP and SHP. The parametric bootstrap mean squared error estimates use 10^4 bootstrap samples. The two mean squared error estimates are virtually identical. Again our posterior variances under the multi-goal prior are much closer to the mean squared error estimates than the corresponding posterior variances under the superharmonic prior.

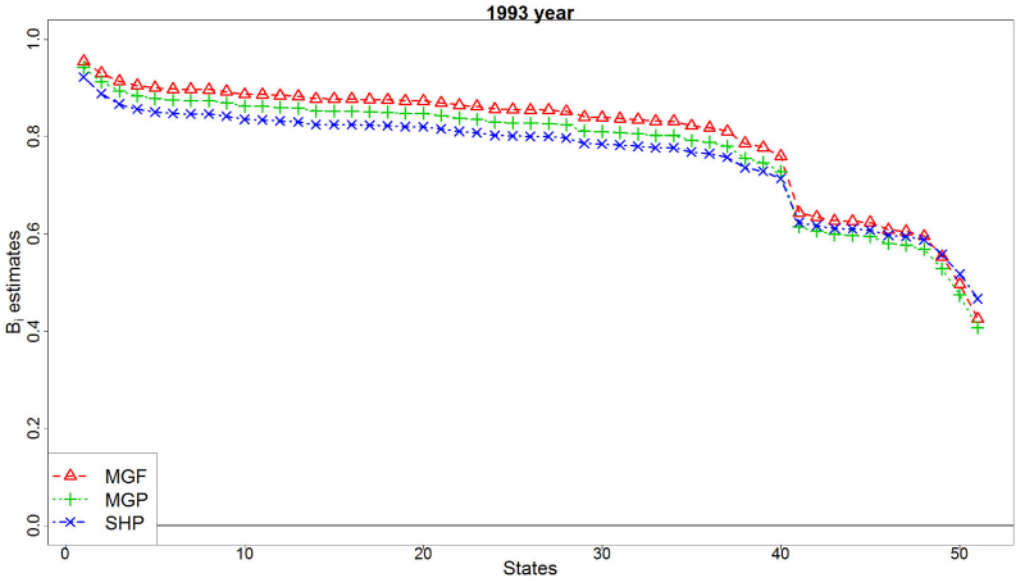

Figure 1. Shrinkage parameter B_i estimates for all the states using three estimation methods, arranged in decreasing order of $\hat{B}_{i;\mathrm{MG}}$; MGF, MGP, and SHP indicate the multi-goal classical estimates $\hat{B}_{i;\mathrm{MG}}$ and Bayes estimates of B_i under the superharmonic and the multi-goal priors, respectively.

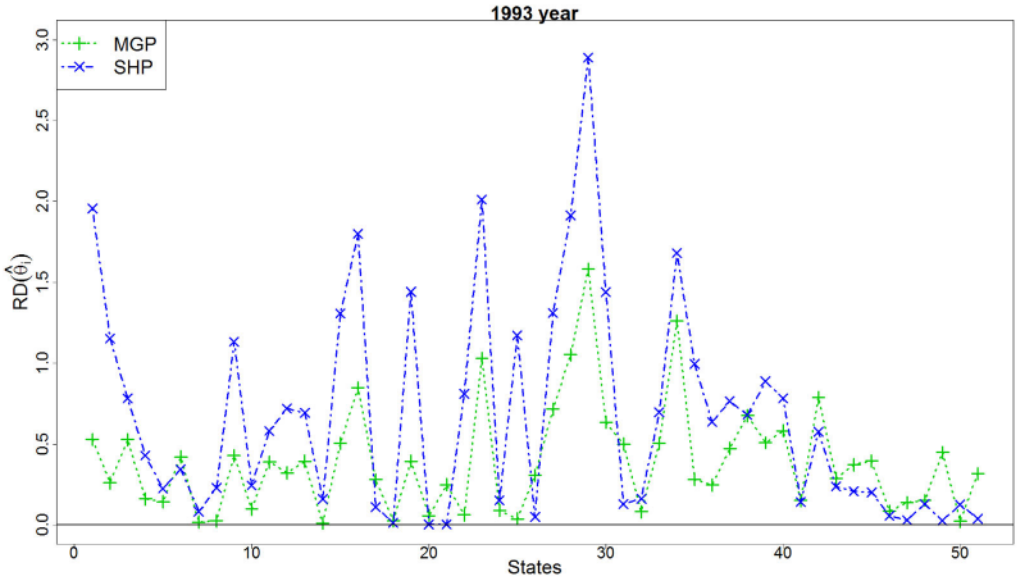

Figure 2. Relative difference $\mathrm{RD}(\hat{\theta}_i)$ of the Bayes estimates $\hat{\theta}_i$ from the corresponding multi-goal classical estimates $\hat{\theta}_{i;\mathrm{MG}}$ for all the states, arranged in decreasing order of $\hat{B}_{i;\mathrm{MG}}$; MGP and SHP indicate two Bayes estimates of θ_i under the multi-goal and the superharmonic priors, respectively.

Next, we present the results using different auxiliary variables for the same SAIPE data. We select one auxiliary variable x_4 with a dummy variable for intercept, described in Section 4 in Hirose and Lahiri (2018). This auxiliary variable provides a moderate coefficient of determination whereas the first setting with all four variables plus the intercept leads to large coefficient of determination. Figures 4–6 display estimates of B_i, relative difference $\mathrm{RD}(\hat{\theta}_i)$ and mean squared error estimates, respectively. Figures 4 and 6 seem to get more closer results between the multi-goal prior (MGP) and the classical estimates (MGF), rather than respective results with all four auxiliary variables. It is also seen that relative difference while using our multi-goal priors still provides the closer results than that under the superharmonic prior.

5. A Discussion on Model Extension

Can we extend our results to a general linear mixed model? To answer this question, we consider the following nested error regression model considered by Battese, Harter, and Fuller (1988):

$$y_{ij} = \theta_{ij} + e_{ij} = x_{ij}'\beta + v_i + e_{ij}, \quad (i = 1, \ldots, m;\ j = 1, \ldots, n_i), \tag{13}$$

where $\{v_1 \ldots, v_m\}$ and $\{e_1, \ldots, e_m\}$ are independent with $v_i \sim N(0, \sigma_v^2)$ and $e_i \sim N(0, \sigma_e^2)$; x_{ij} is a p-dimensional vector of known auxiliary variables; $\beta \in \mathcal{R}^p$ is a p-dimensional vector of unknown regression coefficients; $\psi = (\sigma_v^2, \sigma_e^2)'$ is an unknown variance component vector; n_i is the number of observed unit level data in ith area.

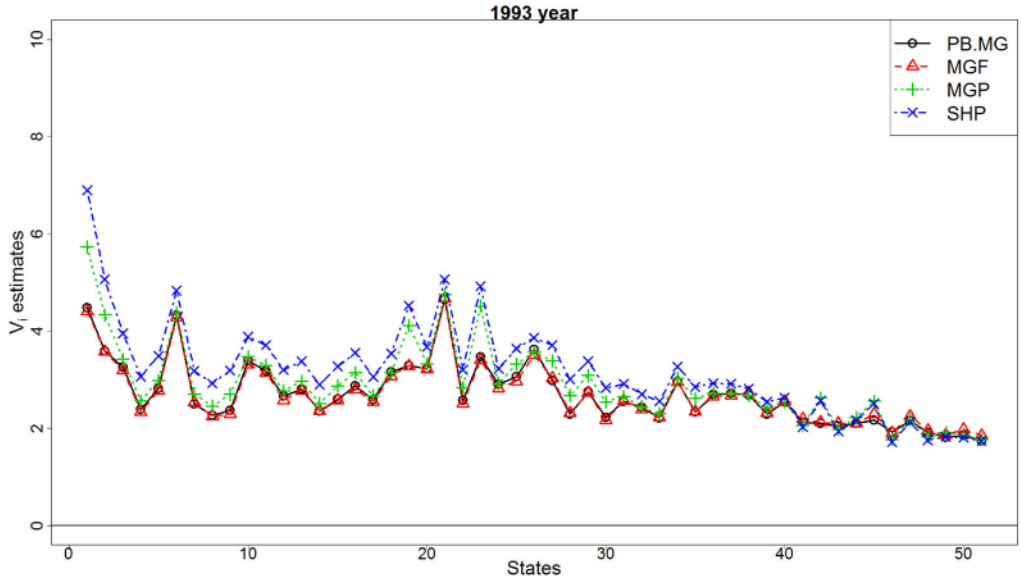

Figure 3. Estimates of the mean squared error of $\hat{\theta}_i$ for all the states using four estimation methods, arranged in decreasing order of $\hat{B}_{i;\mathrm{MG}}$ (PB.MG: mean squared error estimates by parametric bootstrap method $\hat{M}^{*}_{i;\mathrm{MG}}$; MGF: mean squared error estimates by Taylor series method $\hat{M}_{i;\mathrm{MG}}$; MGP: posterior variance under the multi-goal priors; SHP: posterior variance under the superharmonic prior).

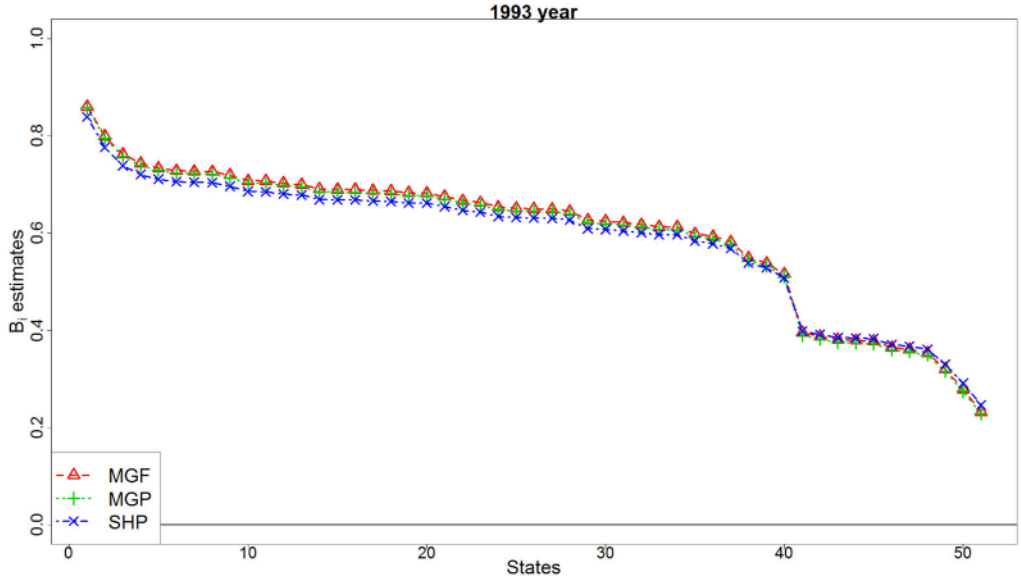

Figure 4. Shrinkage parameter B_i estimates for all the states using three estimation methods using two variables (one auxiliary variable and a dummy variable for intercept), arranged in decreasing order of $\hat{B}_{i;\mathrm{MG}}$; MGF, MGP, and SHP indicate the multi-goal classical estimates $\hat{B}_{i;\mathrm{MG}}$ and Bayes estimates of B_i under the superharmonic and the multi-goal priors, respectively.

The condition for achieving desired property [1] given in Section 3, we need to solve the following system of differential equations with shrinkage factor $B_i = \sigma_e^2/(n_i\sigma_v^2 + \sigma_e^2)$, under certain regularity conditions:

$$\left[\frac{\partial \log h_{i;G}(\psi)}{\partial \psi}\right]' I_F^{-1} \left[\frac{\partial B_i(\psi)}{\partial \psi}\right] = H(\psi), \qquad (14)$$

where

$$\frac{\partial \log h_{i;G}(\psi)}{\partial \psi} = \left(\frac{\partial \log h_{i;G}(\psi)}{\partial \sigma_v^2}, \frac{\partial \log h_{i;G}(\psi)}{\partial \sigma_e^2}\right)',$$

$$H(\psi) = -\frac{1}{2}\mathrm{tr}\left[\frac{\partial^2 B_i(\psi)}{\partial \psi^2} I_F^{-1}\right],$$

$$\frac{\partial B_i(\psi)}{\partial \psi} = \frac{n_i}{(n_i\sigma_v^2 + \sigma_e^2)^2}(-\sigma_e^2, \sigma_v^2)',$$

$$I_F^{-1} = \frac{2}{a}\begin{pmatrix} \sum[(n_i-1)/\sigma_e^4 + (n_i\sigma_v^2 + \sigma_e^2)^{-2}] & -\sum n_i/(n_i\sigma_v^2 + \sigma_e^2)^2 \\ -\sum n_i/(n_i\sigma_v^2 + \sigma_e^2)^2 & \sum n_i^2/(n_i\sigma_v^2 + \sigma_e^2)^2 \end{pmatrix},$$

$$a = \left[\sum n_i^2/(n_i\sigma_v^2 + \sigma_e^2)^2\right]\left[\sum\{(n_i-1)/\sigma_e^4 + (n_i\sigma_v^2 + \sigma_e^2)^{-2}\}\right] - \left[\sum n_i/(n_i\sigma_v^2 + \sigma_e^2)^2\right]^2.$$

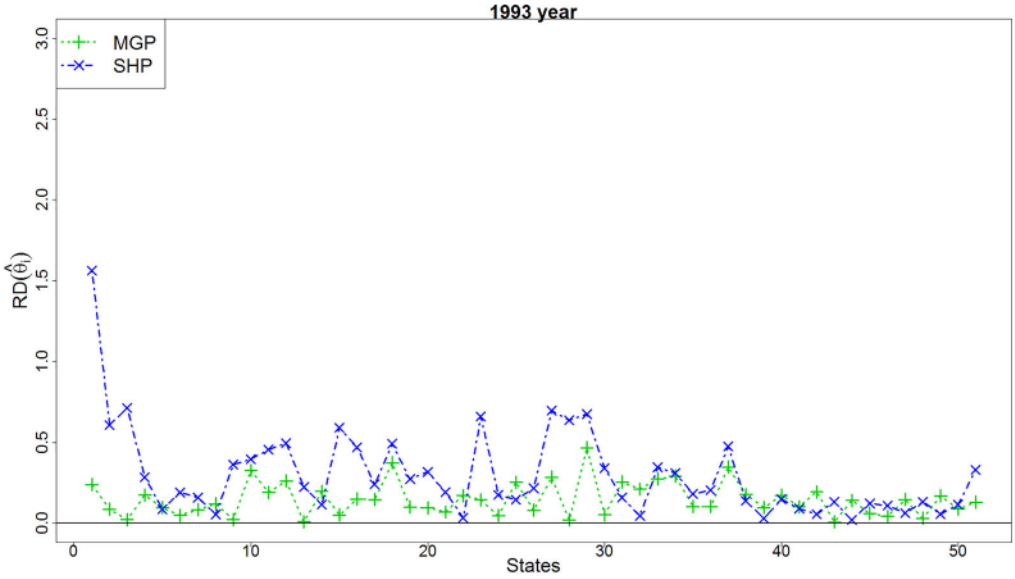

Figure 5. Relative differences RD($\hat{\theta}_i$) of the Bayes estimates $\hat{\theta}_i$ from the corresponding multi-goal classical estimates $\hat{\theta}_{i;\mathrm{MG}}$ for all the states using two variables (one auxiliary variable and a dummy variable for intercept), arranged in decreasing order of $\hat{B}_{i;\mathrm{MG}}$; MGP and SHP indicate two Bayes estimates of θ_i under the multi-goal and the superharmonic priors, respectively.

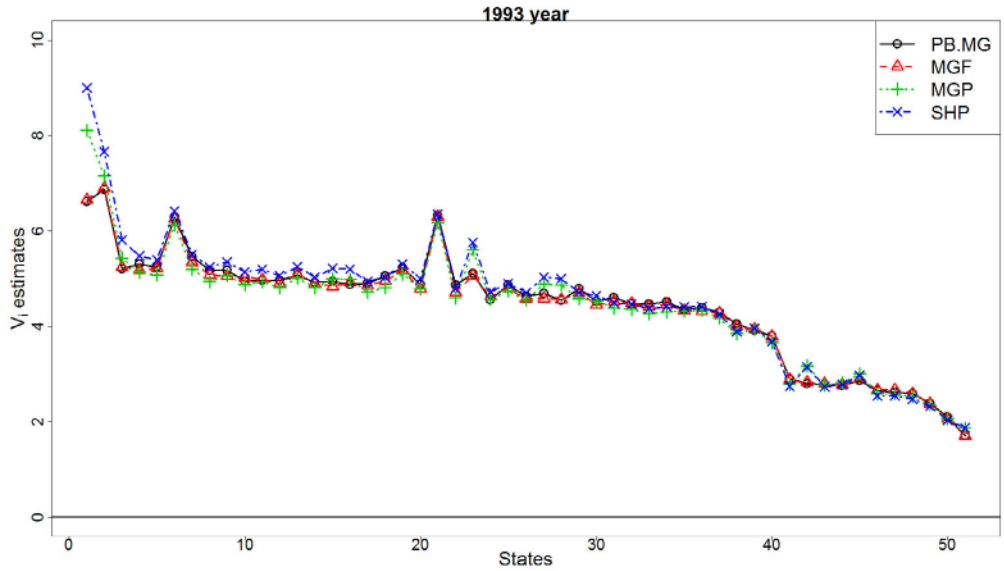

Figure 6. Estimates of the mean squared error of $\hat{\theta}_i$ for all the states using four estimation methods and two variables (one auxiliary variable and a dummy variable for intercept), arranged in decreasing order of $\hat{B}_{i;\mathrm{MG}}$ (PB.MG: mean squared error estimates by parametric bootstrap method $\hat{M}^*_{i;\mathrm{MG}}$; MGF: mean squared error estimates by Taylor series method $\hat{M}_{i;\mathrm{MG}}$; MGP: posterior variance under the multi-goal priors; SHP: posterior variance under the superharmonic prior).

If we use the following adjustment factor $h_{i;G}(\psi)$ for achieving desired property [1]:

$$\frac{\partial \log h_{i;G}(\psi)}{\partial \psi} = \nu k, \tag{15}$$

for a given two dimensional fixed vector k, the solution of ν can be obtained as

$$\nu = \frac{H(\psi)}{k' I_F^{-1} \frac{\partial B_i(\psi)}{\partial \psi}}.$$

This solution thus leads to an appropriate adjustment factor satisfying

$$\frac{\partial \log h_{i;G}(\psi)}{\partial \psi} = \frac{H(\psi)}{k' I_F^{-1} \frac{\partial B_i(\psi)}{\partial \psi}} k.$$

Thus, there exist multiple solutions for $h_{i;G}(\psi)$ satisfying desired property [1] under the nested error regression model (13). Further research is needed to identify a reasonable adjustment factor for the general linear mixed model and to establish a connection with the corresponding Bayesian approach.

Appendix A

We assume the regularity conditions throughout this article as follows:

Regularity conditions:

R1: $\mathrm{rank}(X) = p$ is bounded for large m;

R2: The elements of X are uniformly bounded implying $\sup_{j\geq 1} x_j(X'X)^{-1}x_j = O(m^{-1})$;

R3: $0 < \inf_{i\geq 1} D_i \leq \sup_{i\geq 1} D_i < \infty, A \in (0,\infty)$;

R4: $|\hat{A}_i| < C_{ad} m^{\lambda}$, where $\hat{A}_i$ is an estimator of A and C_{ad} a generic positive constant and λ is small positive constant.

We also restrict the class of adjustment factors $h_+(A)$ and $h_{i;G}(A)$ that satisfy the following regularity conditions, as in Hirose and Lahiri (2018):

R5: $\log h_{i;G}(A)$ is free of y and four times continuously differentiable with respect to A. Moreover, $\frac{\partial^k \log h_{i;G}(A)}{\partial A^k}$ is of order $O(1)$, respectively, for large m with $k = 0, 1, 2, 3$;

R6: $\log h_+(A)$ is free of y and four times continuously differentiable with respect to A. Moreover, $\frac{\partial^k \log h_+(A)}{\partial A^k}$ is of order $o(1)$, for large m with $k = 0, 1, 2, 3$;

R7: $h_+(A)$ is a strictly positive on $A > 0$ satisfying that $h_+(A)\Big|_{A=0} = 0$ and $h_+(A) < C$ on $A > 0$ with a generic positive constant C.

A.1. Proof of Theorem 1

The result follows from an argument similar to the ones given in Das, Jiang, and Rao (2004). We note that for the general adjusted maximum likelihood method (1),

$$l^{(1)}_{i;G}(\hat{A}_{i;G}) - l^{(1)}_{i;G}(A) = (\hat{A}_{i;G} - A)E[l^{(2)}_{i;G}(A)] + (\hat{A}_{i;G} - A)\{l^{(2)}_{i;G}(A) - E[l^{(2)}_{i;G}(A)]\} + \frac{1}{2}(\hat{A}_{i;G} - A)^2 l^{(3)}_{i;G}(A^*_i), \quad \text{(A.1)}$$

where $l^{(k)}_{i;G}(A) = \frac{\partial^k [\tilde{l}_{i;G}(A) + l_{\mathrm{RE}}(A)]}{\partial A^k}$ for $k = 1, 2, 3$ with $\tilde{l}_{i;G}(A) = \log h_{i;G}(A)$ and $\tilde{l}_{\mathrm{RE}}(A) = \log L_{\mathrm{RE}}(A)$. In addition, A^*_i lies between A and $\hat{A}_{i;G}$.

Under regularity conditions, using results of Hirose and Lahiri (2018) and $l^{(1)}_{i;G}(\hat{A}_{i;G}) = 0$, we have $\hat{A}_{i;G} - A = O_p(m^{-1/2})$, $\hat{A}^*_i - A = O_p(m^{-1/2})$, $l^{(1)}_{\mathrm{RE}}(\hat{A}_{i;G}) = -\tilde{l}^{(1)}_{i;G}(\hat{A}_{i;G})$, $E[l^{(2)}_{i;G}(A)] = E[l^{(2)}_{\mathrm{RE}}(A)] + O(1) = -\frac{\mathrm{tr}[V^{-2}]}{2} + O(1)$, $|l^{(2)}_{\mathrm{RE}}(A)| = O_p(m)$, $|l^{(3)}_{\mathrm{RE}}(A)| = O_p(m)$.

Hence, (A.1) yields

$$\begin{aligned}\hat{A}_{i;G} - \hat{A}_{\mathrm{RE}} &= \hat{A}_{i;G} - A - (\hat{A}_{\mathrm{RE}} - A)\\ &= \frac{2}{\mathrm{tr}[V^{-2}]}\tilde{l}^{(1)}_{i;G} + \left\{\frac{2}{\mathrm{tr}[V^{-2}]}\right\}^2 \tilde{l}^{(1)}_{i;G}(A)\{l^{(2)}_{\mathrm{RE}}(A) - E[l^{(2)}_{\mathrm{RE}}(A)]\} + \frac{1}{2}\left\{\frac{2}{\mathrm{tr}[V^{-2}]}\right\}^3 \{\tilde{l}^{(1)}_{i;G}(A)(\tilde{l}^{(1)}_{i;G}(A) + 2l^{(1)}_{\mathrm{RE}}(A))\}\{l^{(3)}_{i;G}(A) + o_p(m)\}. \quad \text{(A.2)}\end{aligned}$$

Using the fact that $l^{(1)}_{\mathrm{RE}}(A) = o_p(m)$,

$$\text{(A.2)} = \frac{2}{\mathrm{tr}[V^{-2}]}\tilde{l}^{(1)}_{i;G} + o_p(m^{-1}). \quad \text{(A.3)}$$

Theorem 1 thus follows.

A.2. Proof of Theorem 2

Proof of part (i). Using Theorem 1, we have

$$\hat{B}_i(\hat{A}_{i;G}) = \hat{B}_i(\hat{A}_{\mathrm{RE}}) - \tilde{l}^{(1)}_{i;G}(A)\frac{2B_i^2}{\mathrm{tr}[V^{-2}]D_i} + o_p(m^{-1}). \quad \text{(A.4)}$$

Hence, using (6) given in (21) of Datta, Rao, and Smith (2005), Equation (2) implies that the following condition is required in order to satisfy $\hat{B}^{\mathrm{HB}}_i = \hat{B}_i(\hat{A}_{i;G})$:

$$\frac{1}{2m\hat{h}_2}\left(\hat{b}_2 - \frac{\hat{h}_3}{\hat{h}_2}\hat{b}_1\right) + \frac{\hat{b}_1}{m\hat{h}_2}\hat{\rho}_1 = -\tilde{l}^{(1)}_{i;G}(A)\frac{2B_i^2}{\mathrm{tr}[V^{-2}]D_i}. \quad \text{(A.5)}$$

Equation (A.5) reduces to

$$\frac{\partial \log \pi_{i;G}(A)}{\partial A} = \tilde{l}^{(1)}_{i;G}(A) + \frac{1}{A + D_i} - \frac{2\mathrm{tr}[V^{-3}]}{\mathrm{tr}[V^{-2}]} + o_p(m^{-1}). \quad \text{(A.6)}$$

After solving the above differential equation, up to the order of $O_p(m^{-1})$, we obtain: $\pi_{i;G}(A) \propto h_{i;G}(A)(A + D_i)\mathrm{tr}[V^{-2}]$.

Part (i) follows from this result. □

Proof of part (ii). Under regularity conditions, Hirose and Lahiri (2018) proved the following result:

$$\mathrm{var}(\hat{B}_i(\hat{A}_{i;G})) = \frac{2D_i^2}{\mathrm{tr}[V^{-2}](A + D_i)^4} + o(m^{-1}).$$

Hence, using the result of Datta, Rao, and Smith (2005),

$$\begin{aligned}V(B_i|y) &= \frac{\hat{b}_1^2}{m\hat{h}_2} + o_p(m^{-1})\\ &= \frac{2D_i^2}{\mathrm{tr}[V^{-2}](A + D_i)^4} + o_p(m^{-1})\\ &= \mathrm{var}(\hat{B}_i(\hat{A}_{i;G})) + o_p(m^{-1}). \quad \text{(A.7)}\end{aligned}$$

Thus, the prior (3) satisfies property (ii) from (A.7).

□

Proof of Part (iii). Datta, Rao, and Smith (2005) obtain the following result:

$$\begin{aligned}E[g_{1i}(A)|y] &= g_{1i}(\hat{A}_{\mathrm{RE}}) + g_{1\pi i}(\hat{A}_{\mathrm{RE}}) + o_p(m^{-1});\\ \theta^{\mathrm{HB}}_i &= y_i - \hat{B}_i(\hat{A}_{\mathrm{RE}})\{y_i - x_i'\hat{\beta}(\hat{A}_{\mathrm{RE}})\}\\ &\quad + \frac{g_{1\pi i}(\hat{A}_{\mathrm{RE}})}{D_i}\{y_i - x_i'\hat{\beta}(\hat{A}_{\mathrm{RE}})\} + o_p(m^{-1}), \quad \text{(A.8)}\end{aligned}$$

where

$$g_{1\pi i}(\hat{A}_{\mathrm{RE}}) = \frac{B_i^2}{m\hat{h}_2}\left(\hat{\rho}_1 - \frac{1}{\hat{A}_{\mathrm{RE}} + D_i} - \frac{\hat{h}_3}{2\hat{h}_2}\right). \quad \text{(A.9)}$$

Using (A.5), we obtain

$$\begin{aligned}g_{1\pi}(\hat{A}_{\mathrm{RE}}) &= \frac{B_i^2}{m\hat{h}_2}\tilde{l}^{(1)}_{i;G}(A) + o_p(m^{-1})\\ &= \frac{2B_i^2}{\mathrm{tr}[V^{-2}]}\tilde{l}^{(1)}_{i;G} + o_p(m^{-1}). \quad \text{(A.10)}\end{aligned}$$

Hence, using Theorem 1, (A.4), (A.8), (A.10) and the fact that $\partial\hat{\beta}(A)/\partial A = O_p(m^{-1/2})$, we have, for large m,

$$\begin{aligned}\theta^{\mathrm{GHB}}_i &= y_i - \hat{B}_i(\hat{A}_{i;G})\{y_i - x_i'\hat{\beta}(\hat{A}_{i;G})\}\\ &\quad + \{\hat{B}_i(\hat{A}_{i;G}) - \hat{B}_i(\hat{A}_{\mathrm{RE}})\}\{y_i - x_i'\hat{\beta}(\hat{A}_{i;G})\}\\ &\quad + \frac{2B_i^2}{\mathrm{tr}[V^{-2}]D_i}\tilde{l}^{(1)}_{i;G}\{y_i - x_i'\hat{\beta}(\hat{A}_{i;G})\} + o_p(m^{-1})\\ &= y_i - \hat{B}_i(\hat{A}_{i;G})\{y_i - x_i'\hat{\beta}(\hat{A}_{i;G})\} + o_p(m^{-1}).\end{aligned}$$

This completes the proof of part (iii). □

Appendix B

B.1. Proof of Remark 1(c)

We show that if we use $h_{i;G}(A)$ alone as a prior, $h_{i;G}(A) = o(A^{(m-p-2)/2})$ is a sufficient condition for the propriety of posterior in a constrained class of adjustment factors $h_{i;G}(A) = (A + D_i)^s$ for some $s > 0$ and fixed m. We note that

$$\begin{aligned}
&\int_0^\infty L_{\mathrm{RE}}(A) h_{i;G}(A) dA \\
&\le C \int_0^\infty (A + \inf_i D_i)^{-m/2} (A + \sup_i D_i)^{p/2+s} dA \\
&= C \int_0^\infty \left[\frac{(A + \sup_i D_i)}{(A + \inf_i D_i)} \right]^{m/2} (A + \sup_i D_i)^{-m/2+p/2+s} dA \\
&\le C \int_{\sup_i D_i}^\infty t^{-m/2+p/2+s} dt. \qquad \text{(B.1)}
\end{aligned}$$

It is evident that the condition $s < (m - p - 2)/2$ achieves (B.1) $< \infty$. Thus, the condition $h_{i;G}(A) = o(A^{(m-p-2)/2})$ is a sufficient condition for it to be a bonafide prior for large A.

The following inequality shows that $\pi_{i;G}(A)$ could be a prior if the condition $h_{i;G}(A) = o(A^{(m-p)/2})$ is met.

$$\begin{aligned}
&\int_0^\infty L_{\mathrm{RE}}(A) \pi_{i;G}(A) dA \\
&\le C \int_0^\infty (A + \inf_i D_i)^{-m/2-2} (A + \sup_i D_i)^{p/2+1+s} dA \\
&= C \int_0^\infty \left[\frac{(A + \sup_i D_i)}{(A + \inf_i D_i)} \right]^{m/2+2} (A + \sup_i D_i)^{-m/2-2+p/2+1+s} dA \\
&\le C \int_{\sup_i D_i}^\infty t^{-m/2+p/2-1+s} dt. \qquad \text{(B.2)}
\end{aligned}$$

Hence, if $h_{i;G}(A)$ in $\pi_{i;G}(A)$ satisfies $s < (m - p)/2$, then we have (B.2) $< \infty$. Thus, the condition $h_{i;G}(A) = o(A^{(m-p)/2})$ is a sufficient condition for $\pi_{i;G}(A)$ being a bonafide prior in a Bayesian method as well as an adjustment factor in an adjusted maximum likelihood method.

B.2. Proof of Remark 1(d)

The following result follows from (A.3) and the regularity condition R5 for all i,

$$\hat{A}_{i;G} - A - (\hat{A}_{\mathrm{RE}} - A) = O(m^{-1}) + o_p(m^{-1}).$$

Moreover, under the regularity conditions, the following results follow from Yoshimori and Lahiri (2014a) and Hirose and Lahiri (2018).

$$\begin{aligned}
E[A_{\mathrm{RE}} - A] &= o(m^{-1}), \\
E[A_{i;G} - A] &= O(m^{-1}), \\
E[(A_{\mathrm{RE}} - A)^2] &= \frac{2}{\mathrm{tr}[V^{-2}]} + o(m^{-1}), \\
E[(A_{i;G} - A)^2] &= \frac{2}{\mathrm{tr}[V^{-2}]} + o(m^{-1}).
\end{aligned}$$

Hence, we have for large m and all i,

$$\begin{aligned}
&\frac{\hat{A}_{i;G} - E[\hat{A}_{i;G}]}{\sqrt{V[\hat{A}_{i;G}]}} - \frac{\hat{A}_{\mathrm{RE}} - E[\hat{A}_{\mathrm{RE}}]}{\sqrt{V[\hat{A}_{\mathrm{RE}}]}} \\
&= \frac{\hat{A}_{i;G} - A}{\sqrt{2/\mathrm{tr}[V^{-2}]}} - \frac{\hat{A}_{\mathrm{RE}} - A}{\sqrt{2/\mathrm{tr}[V^{-2}]}} + O(m^{-1/2}), \\
&= O(m^{-1/2}) + o_p(m^{-1/2}).
\end{aligned}$$

Each estimator $\hat{A}_{i;G}$, therefore, has the same asymptotic distribution as that of $\hat{A}_{\mathrm{RE}}$ for large m under the regularity conditions. This also implies that the estimator $\hat{A}_{i;G}$ has asymptotic normality and efficiency as well as the asymptotic properties of the residual maximum likelihood estimator $\hat{A}_{\mathrm{RE}}$ given in Jiang (1996).

We next show that the covariance matrix of $(\hat{A}_{1;G}, \ldots, \hat{A}_{m;G})$ converges to a singular matrix for large m. As for the component of the covariance matrix of $(\hat{A}_{1;G}, \ldots, \hat{A}_{m;G})$, we obtain for any i and j such that $i \neq j$,

$$\begin{aligned}
\mathrm{cov}(\hat{A}_{i;G}, \hat{A}_{j;G}) &= E[(\hat{A}_{i;G} - E[\hat{A}_{i;G}])(\hat{A}_{j;G} - E[\hat{A}_{j;G}])], \\
&= E[(\hat{A}_{i;G} - A - E[(\hat{A}_{i;G} - A)]) \\
&\quad (\hat{A}_{j;G} - A - E[(\hat{A}_{j;G} - A)])], \\
&= \frac{4}{\mathrm{tr}[V^{-2}]^2} E\left[(l_{\mathrm{RE}}^{(1)} + l_{i;G}^{(1)})(l_{\mathrm{RE}}^{(1)} + l_{j;G}^{(1)}) \right] + o(m^{-1}), \\
&= \frac{4}{\mathrm{tr}[V^{-2}]^2} E\left[(l_{\mathrm{RE}}^{(1)})^2 \right] + o(m^{-1}), \\
&= \frac{2}{\mathrm{tr}[V^{-2}]} + o(m^{-1}), \\
V(\hat{A}_{i;G}) &= \frac{2}{\mathrm{tr}[V^{-2}]} + o(m^{-1}).
\end{aligned}$$

Note that we use the following results in the above calculations.

$$\begin{aligned}
E[(l_{\mathrm{RE}}^{(1)})^2] &= \frac{1}{4} E[(y' P^2 y - \mathrm{tr}[P])^2] \\
&= \frac{\mathrm{tr}[P^2]}{2} = \frac{\mathrm{tr}[V^{-1}]}{2} + O(1), \\
E[(\hat{A}_{i;G} - A)(\hat{A}_{j;G} - A)] &= \frac{4}{\mathrm{tr}[V^{-2}]^2} E\left[(l_{\mathrm{RE}}^{(1)} + l_{i;G}^{(1)})(l_{\mathrm{RE}}^{(1)} + l_{j;G}^{(1)}) \right] \\
&\quad + o(m^{-1}).
\end{aligned}$$

The later results can be shown using a calculation similar to that of Theorem 4 of Das, Jiang, and Rao (2004). The result thus follows.

B.3. Proof of Remark 2(a)

We show that the prior (10) achieves (ii)–(v).

Proof of (ii). From the results of Datta, Rao, and Smith (2005) and Hirose and Lahiri (2018),

$$\begin{aligned}
V(B_i|y) &= \frac{\hat{b}_1^2}{m\hat{h}_2} + o_p(m^{-1}) \\
&= \frac{2D_i^2}{\mathrm{tr}[V^{-2}](A + D_i)^4} + o_p(m^{-1}) \\
&= \mathrm{var}(\hat{B}_{i;\mathrm{MG}}) + o_p(m^{-1}). \qquad \text{(B.3)}
\end{aligned}$$

Hence, the prior achieves the property (ii) from (B.3). □

Proof of part (iii). Using (5), it is straightforward to show

$$g_{1i}(\hat{A}_{i;\mathrm{MG}}) - g_{1i}(\hat{A}_{\mathrm{RE}}) = \frac{2D_i^2}{\mathrm{tr}[V^{-2}](A + D_i)^3} + o_p(m^{-1}).$$

Using (7) and (A.9), we obtain the following after some algebra:

$$g_{1i}(\hat{A}_{i;\mathrm{MG}}) = g_{1i}(\hat{A}_{\mathrm{RE}}) + g_{1\pi i}(\hat{A}_{\mathrm{RE}}) + o_p(m^{-1}). \qquad \text{(B.4)}$$

Using (A.8), Corollary 1 (ii) and (B.4), we get

$$\begin{aligned}
\theta_i^{\mathrm{HB}} &= y_i - \hat{B}_i(\hat{A}_{i;\mathrm{MG}})\{y_i - x_i'\hat{\beta}(\hat{A}_{i;\mathrm{MG}})\} \\
&\quad + \{\hat{B}_i(\hat{A}_{i;\mathrm{MG}}) - \hat{B}_i(\hat{A}_{\mathrm{RE}})\}\{y_i - x_i'\hat{\beta}(\hat{A}_{i;\mathrm{MG}})\} \\
&\quad + \{\hat{B}_i(\hat{A}_{\mathrm{RE}}) - \hat{B}_i(\hat{A}_{i;\mathrm{MG}})\}\{y_i - x_i'\hat{\beta}(\hat{A}_{i;\mathrm{MG}})\} + o_p(m^{-1}) \\
&= \hat{\theta}_{i;\mathrm{MG}}^{\mathrm{EB}} + o_p(m^{-1}). \qquad \text{(B.5)}
\end{aligned}$$

Property (iii) thus follows from the result (B.5). □

Proof of parts (iv)–(v). Using (B.4), we get

$$E[g_{1i}(A)|y] = g_{1i}(\hat{A}_{i;\mathrm{MG}}) + o_p(m^{-1}). \tag{B.6}$$

Datta, Rao, and Smith (2005) obtained the following result:

$$V[\theta_i|y] = g_{1i}(\hat{A}_{\mathrm{RE}}) + g_{1\pi i}(\hat{A}_{\mathrm{RE}}) + g_{2i}(\hat{A}_{\mathrm{RE}}) + g_{4i}(\hat{A}_{\mathrm{RE}}; y_i) + o_p(m^{-1}). \tag{B.7}$$

Using the result given in Butar and Lahiri (2003), Hirose and Lahiri (2018), (B.4) and (B.6), we get

$$\begin{aligned} V[\theta_i|y] =& g_{1i}(\hat{A}_{i;\mathrm{MG}}) + g_{2i}(\hat{A}_{i;\mathrm{MG}}) + g_{3i}(\hat{A}_{i;\mathrm{MG}}) + o_p(m^{-1}) \\ =& \hat{M}_i(\hat{A}_{i;\mathrm{MG}}) + o_p(m^{-1}) \\ =& M_i(\hat{\theta}^{\mathrm{EB}}_{i;\mathrm{MG}}) + o_p(m^{-1}) \\ =& \hat{M}^{\mathrm{boot}}_{i;\mathrm{MG}} + o_p(m^{-1}). \end{aligned} \tag{B.8}$$

Equation (B.8) implies that the prior (10) also satisfies (iv)–(v) simultaneously. □

Acknowledgments

We thank Professor Shuhei Mano, anonymous associate editor, and referees for reading an earlier version of the article carefully and offering a number of constructive suggestions, which led to a significant improvement of our article.

Funding

The first and second authors' research was partially supported by JSPS KAKENHI grant number 18K12758 and U.S. National Science Foundation grants SES-1534413 and SES-1758808, respectively.

References

Battese, G. E., Harter, R. M., and Fuller, W. A. (1988), "An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data," *Journal of the American Statistical Association*, 83, 28–36. [6]

Bell, W. R., Basel W. W., and Maples, J. J. (2016), "An Overview of the U.S. Census Bureau's Small Area Income and Poverty Estimates Program," in *Analysis of Poverty Data by Small Area Estimation*, ed. M. Pratesi, West Sussex: Wiley, pp. 349–377. [5]

Bell, W. R., and Franco, C. (2017), "Small Area Estimation-State Poverty Rate Model Research Data Files," available at *https://www.census.gov/srd/csrmreports/byyear.html*. [5]

Butar, F. B., and Lahiri, P. (2003), "On Measures of Uncertainty of Empirical Bayes Small-Area Estimators," *Journal of Statistical Planning and Inference*, 112, 63–76. [11]

Carter, G. M., and Rolph, J. F. (1974), "Empirical Bayes Methods Applied to Estimating Fire Alarm Probabilities," *Journal of the American Statistical Association*, 69, 880–885. [1]

Casas-Cordero, C., Encina, J., and Lahiri, P. (2016), "Poverty Mapping for the Chilean Comunas," in *Analysis of Poverty Data by Small Area Estimation*, ed. M. Pratesi, New York: Wiley, pp. 379–403. [1]

Citro, C., and Kalton, G., eds. (2000), *Small-Area Income and Poverty Estimates: Priorities for 2000 and Beyond*, Panel on Estimates of Poverty for Small Geogruphic Area, Washington, DC: National Academy Press. [1]

Das, K., Jiang, J., and Rao, J. N. K. (2004), "Mean Squared Error of Empirical Predictor," *The Annals of Statistics*, 32, 818–840. [3,9,10]

Datta, G. S., Rao, J. N. K., and Smith, D. D. (2005), "On Measuring the Variability of Small Area Estimators Under a Basic Area Level Model," *Biometrika*, 92, 183–196. [2,3,4,5,9,10,11]

Efron, B., and Morris, C. (1973), "Stein's Estimation Rule and Its Competitors an Empirical Bayes Approach," *Journal of the American Statistical Association*, 68, 117–130. [1]

——— (1975), "Data Analysis Using Stein's Estimator and Its Generalizations," *Journal of the American Statistical Association*, 70, 311–319. [1]

Erciulescu, A. L., Franco, C., and Lahiri, P. (2020), "Use of Administrative Records in Small Area Estimation," in *Administrative Records for Survey Methodology*, Wiley Series in Survey Methodology, eds. A. Y. Chun and M. Larsen, forthcoming. [5]

Fay, R. E., and Herriot, R. A. (1979), "Estimates of Income for Small Places: An Application of James–Stein Procedures to Census Data," *Journal of the American Statistical Association*, 74, 269–277. [1]

Ganesh, N., and Lahiri, P. (2008), "A New Class of Average Moment Matching Priors," *Biometrika*, 95, 514–520. [5]

Hirose, M. Y. (2017), "Non-Area-Specific Adjustment Factor for Second-Order Efficient Empirical Bayes Confidence Interval," *Computational Statistics & Data Analysis*, 116, 67–78. [2]

——— (2019), "A Class of General Adjusted Maximum Likelihood Methods for Desirable Mean Squared Error Estimation of EBLUP Under the Fay–Herriot Small Area Model," *Journal of Statistical Planning and Inference*, 199, 302–310. [2]

Hirose, M. Y., and Lahiri, P. (2018), "Estimating Variance of Random Effects to Solve Multiple Problems Simultaneously," *The Annals of Statistics*, 46, 1721–1741. [2,4,5,6,9,10,11]

James, W., and Stein, C. (1961), "Estimation With Quadratic Loss," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. I), Berkeley, CA: University of California Press, pp. 361–379. [1]

Jiang, J. (1996), "REML Estimation: Asymptotic Behavior and Related Topics," *The Annals of Statistics*, 24, 255–286. [10]

——— (2007), *Linear and Generalized Linear Mixed Models and Their Applications*, New York: Springer. [1]

Lahiri, P. (2003), "A Review of Empirical Best Linear Unbiased Prediction for the Fay-Herriot Small-Area Model," *The Philippine Statistician*, 52, 1–15. [1]

Lahiri, P., and Li, H. (2009), "Generalized Maximum Likelihood Method in Linear Mixed Models With an Application in Small Area Estimation," in *Proceedings of the Federal Committee on Statistical Methodology Research Conference*. [2]

Li, H., and Lahiri, P. (2010), "An Adjusted Maximum Likelihood Method for Solving Small Area Estimation Problems," *Journal of Multivariate Analysis*, 101, 882–892. [2,3]

Mohadjer, L., Rao, J. N. K., Liu, B., Krenzke, T., and Van de Kerckhove, W. (2012), "Hierarchical Bayes Small Area Estimates of Adult Literacy Using Unmatched Sampling and Linking Models," *Journal of the Indian Society of Agricultural Statistics*, 66, 55–63, 232–233. [1]

Morris, C. N. (1983), "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association*, 78, 47–65. [2]

Morris, C., and Tang, R. (2011), "Estimating Random Effects via Adjustment for Density Maximization," *Statistical Science*, 26, 271–287. [2]

Otto, M. C., and Bell, W. R. (1995), "Sampling Error Modeling of Poverty and Income Statistics for States," in *Proceedings of the Section on Government Statistics*, Alexandria, VA: American Statistical Association, pp. 160–165. [1]

Rao, J. N. K., and Molina, I. (2015), *Small Area Estimation* (2nd ed.), New York: Wiley. [1]

Yoshimori, M., and Lahiri, P. (2014a), "A New Adjusted Maximum Likelihood Method for the Fay-Herriot Small Area Model," *Journal of Multivariate Analysis*, 124, 281–294. [2,3,4,10]

——— (2014b), "A Second-Order Efficient Empirical Bayes Confidence Interval," *The Annals of Statistics*, 42, 1–29. [2,3]