Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Bayesian benchmarking of the Fay-Herriot model using random deletion

by Balgobin Nandram, Andreea L. Erciulescu and Nathan B. Cruze

Release date: June 27, 2019

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version **is also available.**

Cette publication est aussi disponible en français.

Survey Methodology, June 2019
Vol. 45, No. 2, pp. 365-390
Statistics Canada, Catalogue No. 12-001-X

Bayesian benchmarking of the Fay-Herriot model using random deletion

Balgobin Nandram, Andreea L. Erciulescu and Nathan B. Cruze[1]

Abstract

Benchmarking lower level estimates to upper level estimates is an important activity at the United States Department of Agriculture's National Agricultural Statistical Service (NASS) (e.g., benchmarking county estimates to state estimates for corn acreage). Assuming that a county is a small area, we use the original Fay-Herriot model to obtain a general Bayesian method to benchmark county estimates to the state estimate (the target). Here the target is assumed known, and the county estimates are obtained subject to the constraint that these estimates must sum to the target. This is an external benchmarking; it is important for official statistics, not just NASS, and it occurs more generally in small area estimation. One can benchmark these estimates by "deleting" one of the counties (typically the last one) to incorporate the benchmarking constraint into the model. However, it is also true that the estimates may change depending on which county is deleted when the constraint is included in the model. Our current contribution is to give each small area a chance to be deleted, and we call this procedure the random deletion benchmarking method. We show empirically that there are differences in the estimates as to which county is deleted and that there are differences of these estimates from those obtained from random deletion as well. Although these differences may be considered small, it is most sensible to use random deletion because it does not give preferential treatment to any county and it can provide small improvement in precision over deleting the last one benchmarking as well.

Key Words: Constraint; Direct estimates; Fay-Herriot model; Multivariate normal density; Official statistics; Small area estimation.

1 Introduction

In official statistics, it is important for lower level estimates to sum to upper level estimates. For example, the National Agricultural Statistics Service (NASS) often uses a "top-down" sequence in the release of its official estimates in which national and state estimates, e.g., estimated corn acreage totals, are published prior to the completion of supplemental data collection and estimation of corresponding county estimates (Cruze, Erciulescu, Nandram, Barboza and Young, 2019). Within these small administrative areas, the survey data often become sparse. Several popular modeling techniques give rise to more reliable small area estimates. However, the small area estimates may not automatically satisfy relationships with estimates at other levels of aggregation, and benchmarking procedures may be applied to enforce consistency among estimates.

There is a considerable history on benchmarking techniques which have been used to impose agreement among multiple levels and to protect against possible model misspecification. These procedures can be broadly classified in two categories: internal benchmarking, in which a target is derived from current survey data, and external benchmarking, in which a desired target may be taken from other sources such as administrative data or previously established estimates. We discuss external benchmarking, in accordance with NASS's "top-down" procedure, of the Fay-Herriot (FH) model (Fay and Herriot, 1979).

1. Balgobin Nandram, Worcester Polytechnic Institute and USDA National Agricultural Statistics Service, Department of Mathematical Sciences, Stratton Hall, 100 Institute Road, Worcester, MA 01609. E-mail: balnan@wpi.edu; Andreea L. Erciulescu, Westat, 1600 Research Boulevard, Rockville, MD 20850. E-mail: alerciulescu@gmail.com; Nathan B. Cruze, USDA National Agricultural Statistics Service, 1400 Independence Avenue, SW, Room 6412 A, Washington, DC 20250-2054. E-mail: nathan.cruze@nass.usda.gov.

The most recent review of small area estimation is given in Pfeffermann (2013), but see Rao and Molina (2015) for the most updated textbook on small area estimation. Earlier Jiang and Lahiri (2006) gave an extensive review of the classical inferential approach for linear and generalized linear mixed models that are used in small area estimation. There are discussions of benchmarking in these works as well, but the latter review was not on the hierarchical Bayes approach that is of primary interest in this paper.

Within the hierarchical Bayes framework, You, Rao and Dick (2004) studied benchmarked estimators for small area estimation based on unmatched sampling and linking models proposed earlier by You and Rao (2002). They applied this approach to undercoverage estimation for the ten provinces across Canada for the 1991 Canadian Census. Wang, Fuller and Qu (2008) gave a characterization of the best linear unbiased predictor (BLUP) for small area means under an area level model that satisfies a benchmarking constraint and minimizes the loss function criterion that all linear unbiased predictors satisfy. They also presented an alternative way of imposing the benchmarking constraint such that the BLUP estimator would have a self-calibrated property (discussed in You and Rao, 2002). Wang et al. (2008) characterized a class of benchmarked estimators as the predictors that minimize a quadratic loss function subject to a benchmarking restriction. Their proposed self-calibrated augmented model reduces bias both at the overall and small area level. Other benchmarking procedures are given by Bell, Datta and Ghosh (2013), Ghosh and Steorts (2013), Pfeffermann, Sikov and Tiller (2014) and Pfeffermann and Tiller (2006).

Whether fitting unit-level or area-level models, incorporating a fixed, external target amounts to imposing the general constraint $\sum_{i=1}^{\ell} w_i \theta_i = a$, where a is a known constant and the θ_i denote small area estimates to be benchmarked; for totals, the weights w_i are all equal to 1. One way to do so is by using the following transformation, $\phi = a - \sum_{i=1}^{\ell} \theta_i$, keeping θ_i, $i = 1, \ldots, \ell - 1$, unchanged and "deleting" the last small area, replacing it with $\theta_\ell = \phi - \left(a - \sum_{i=1}^{\ell-1} \theta_i\right)$. Janicki and Vesper (2017) introduced a slightly different transformation, $\phi_i = \theta_i$, $i = 1, \ldots, \ell - 1$, $\phi_\ell = \sum_{i=1}^{\ell} \theta_i$, which is essentially an internal benchmarking that preserves the sum of all ℓ estimates. If that sum (of all ℓ unbenchmarked estimates) were prescribed as an external target, then $\theta_\ell = \phi_\ell - \sum_{i=1}^{\ell-1} \theta_i$, and Janicki and Vesper's transformation becomes equivalent to deleting last small area estimate.

External benchmarking procedures, which deleted the last small area estimate, were explored by Nandram and Sayit (2011) and by Nandram, Berg, and Barboza (2014) for the purposes of benchmarking binomial probabilities and forecasts of crop yield, respectively. (In both of these contexts the constraint was actually imposed on the weighted sum of small area estimates.) Erciulescu, Cruze, and Nandram (2019) considered a variety of external benchmarking techniques including deletion, difference benchmarking, and ratio benchmarking in the context of hierarchical Bayesian small area models. Collectively, the external benchmarking constraint has been inserted in the likelihood function (e.g., Toto and Nandram, 2010), the joint density of the area effects (e.g., Nandram and Sayit, 2011), or in the posterior density of the area effects (Janicki and Vesper, 2017), although the latter choice is using the prior knowledge or requirements embodied in the constraint a posteriori rather than on the prior distributions themselves.

Datta, Ghosh, Steorts and Maples (2011), henceforth DGSM, proposed a general class of constrained Bayes estimators to provide benchmarked estimates. Referring specifically to the method of Toto and

Nandram (2010) for unit level models, DGSM wrote the following: *"A disadvantage to such an approach is that results can differ depending on which unit is dropped"*. This statement also applies to Nandram and Toto (2010), Nandram, Toto and Choi (2011), Janicki and Vesper (2017) and others. It also applies in the same way to an area-level model subject to an external constraint. The procedures of DGSM depend on an important area-specific parameter (see Section 4). This parameter also has several different specifications, and it can be argued that the resulting estimates could also be affected by the choice of specification. Moreover, the procedures of DGSM do not provide posterior standard errors or credible intervals.

In response to DGSM's comment on the last area deletion benchmarking, we introduce a random deletion benchmarking, giving a chance to each area to be deleted, and not just the last one. The random deletion benchmarking method is motivated mathematically in Appendix A. Empirical results show that there are slight differences between the last one deletion benchmarking and the random deletion benchmarking.

In this paper, we discuss random deletion benchmarking in the context of a Bayesian FH (BFH) model. In Section 2, the BFH model without constraint is introduced. The methodology for imposing an external target on the BFH model through random deletion is developed in Section 3. In Section 4, we describe the empirical studies to assess features of estimates obtained from random deletion benchmarking, including related measures of uncertainty. Finally, Section 5 has concluding remarks; more technical details are provided in several appendices.

2 Bayesian Fay-Herriot model

Assume that the observed data are $\left(\hat{\theta}_i,\, s_i\right)$, $i = 1, \ldots, \ell$, where $\hat{\theta}_i$ and s_i are respectively an estimate and its standard error (for simplicity, assumed known) of a quantity under study, e.g., the i^{th} area total θ_i. The BFH model is

$$\hat{\theta}_i \,\big|\, \theta_i \overset{\text{ind}}{\sim} \text{Normal}\left(\theta_i,\, s_i^2\right),$$

$$\theta_i \,|\, \boldsymbol{\beta},\, \sigma^2 \overset{\text{ind}}{\sim} \text{Normal}\left(\mathbf{x}_i'\boldsymbol{\beta},\, \sigma^2\right),$$

$$\pi\left(\boldsymbol{\beta},\, \sigma^2\right), \tag{2.1}$$

where $i = 1, \ldots, \ell$, $\mathbf{x}_i$ are a set of covariates with p components (including intercept) and $\pi\left(\boldsymbol{\beta},\, \sigma^2\right)$ is the joint prior distribution for $\left(\boldsymbol{\beta},\, \delta^2\right)$. A priori it is assumed that $\pi\left(\boldsymbol{\beta},\, \sigma^2\right) = \pi\left(\boldsymbol{\beta}\right)\pi\left(\sigma^2\right)$, i.e., $\boldsymbol{\beta}$ and σ^2 are independent, with

$$\pi\left(\boldsymbol{\beta}\right) \propto 1 \quad \text{and} \quad \pi\left(\sigma^2\right) = 1\big/\left(1+\sigma^2\right)^2. \tag{2.2}$$

By Bayes' theorem, the joint posterior density of $\ell + p + 1$ model parameters is

$$\pi\left(\boldsymbol{\theta},\, \boldsymbol{\beta},\, \sigma^2 \,|\, \hat{\boldsymbol{\theta}}\right) \propto \frac{1}{\left(1+\sigma^2\right)^2}\left(\frac{1}{\sigma^2}\right)^{\ell/2} \prod_{i=1}^{\ell}\left\{\exp\left[-\frac{1}{2}\left\{\frac{1}{s_i^2}\left(\hat{\theta}_i - \theta_i\right)^2 + \frac{1}{\sigma^2}\left(\theta_i - \mathbf{x}_i'\boldsymbol{\beta}\right)^2\right\}\right]\right\}. \tag{2.3}$$

In (2.2), $\pi(\sigma^2)$ is a proper prior distribution, flatter than $\pi(\sigma^2) \propto 1/\sigma^2$ near zero, with no moments. In fact, any proper prior for σ^2 is fine; an improper prior on σ^2 may lead to improper posterior density. Because $\pi(\boldsymbol{\beta})$ is improper, the product $\pi(\boldsymbol{\beta}, \sigma^2) = \pi(\boldsymbol{\beta})\pi(\sigma^2)$ *is improper*, and this could cause the joint posterior density of $(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2)$ to be improper, an undesirable scenario. Theorem 1 below establishes the propriety of the joint posterior density (2.3).

More details about the BFH model are presented in Appendix B. Specifically, letting $\lambda_i = \frac{\sigma^2}{s_i^2+\sigma^2}$, $i = 1, \ldots, \ell$, we have shown that

$$\theta_i \,|\, \boldsymbol{\beta}, \sigma^2, \hat{\boldsymbol{\theta}} \overset{\text{ind}}{\sim} \text{Normal}\left\{\lambda_i\hat{\theta}_i + (1-\lambda_i)\mathbf{x}_i'\boldsymbol{\beta}, (1-\lambda_i)\sigma^2\right\}, \; i = 1, \ldots, \ell, \tag{2.4}$$

$$\boldsymbol{\beta} \,|\, \sigma^2, \hat{\boldsymbol{\theta}} \sim \text{Normal}\left(\hat{\boldsymbol{\beta}}, \hat{\Sigma}\right), \tag{2.5}$$

$$\pi_3\left(\sigma^2 \,|\, \hat{\boldsymbol{\theta}}\right) \propto Q(\sigma^2)\frac{1}{(1+\sigma^2)^2}, \tag{2.6}$$

where

$$\hat{\boldsymbol{\beta}} = \hat{\Sigma}\sum_{i=1}^{\ell}\frac{\hat{\theta}_i\mathbf{x}_i}{s_i^2+\sigma^2}, \; \hat{\Sigma}^{-1} = \sum_{i=1}^{\ell}\frac{\mathbf{x}_i\mathbf{x}_i'}{s_i^2+\sigma^2},$$

$$Q(\sigma^2) = \left|\hat{\Sigma}\right|^{1/2}\prod_{i=1}^{\ell}\frac{1}{(s_i^2+\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{\ell}\frac{1}{s_i^2+\sigma^2}\left(\hat{\theta}_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}\right)^2\right\}.$$

Theorem 1

The joint posterior density (2.3) is proper provided the design matrix is full rank.

Proof of Theorem 1

Because the design matrix is full rank, $\hat{\boldsymbol{\beta}}$ and $\hat{\Sigma}$ are well defined for all σ^2. This implies that $Q(\sigma^2)$ is bounded in σ^2. Therefore, as $\pi(\sigma^2) = \frac{1}{(1+\sigma^2)^2}$ is proper, the posterior density $\pi_3\left(\sigma^2 \,|\, \hat{\boldsymbol{\theta}}\right)$ is proper. Then, by applying the multiplication rule of probability, it follows that the joint posterior density, $\pi\left(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2 \,|\, \hat{\boldsymbol{\theta}}\right)$, is proper.

With propriety assured, sampling from the joint posterior density and inference about θ_i can be achieved through a simple Monte Carlo procedure. Using the multiplication rule and drawing samples from (2.6), (2.5) and (2.4), the procedure follows. First, draw a sample from $\pi_3\left(\sigma^2 \,|\, \hat{\boldsymbol{\theta}}\right)$ in (2.6). Draws from this distribution can be made using a grid method; see Appendix B. It is then easy to draw samples from the conditional posterior density of $\boldsymbol{\beta}$ in (2.5). Finally, samples of the θ_i can be drawn independently from (2.4). Sampling in this manner from the joint posterior density under the unconstrained BFH model does not need monitoring (unlike Markov chain Monte Carlo methods). The unconstrained BFH model will

provide a basis for comparison of the estimates and related measures of uncertainty obtained under the proposed random deletion benchmarking method.

3 Random deletion methodology

As remarked earlier, the random deletion methodology is obtained by introducing a new variable which takes the values $1, \ldots, \ell$ with equal, possibly different, probabilities (weights). In Section 3.1, we show how to construct the joint posterior density of the parameters of the BFH model under random deletion, and in Section 3.2, we show how to sample from the joint posterior density.

3.1 Construction of the joint posterior density

The Basic Benchmarking Theorem is motivated by the work reviewed in Section 1. The next goal is to construct the joint prior density of $\theta_1, \ldots, \theta_\ell$ subject to the benchmarking constraint that $\sum_{i=1}^{\ell} \theta_i = a$, where a is a known external target. The joint prior density of $\theta_1, \ldots, \theta_\ell$ will be used to complete the constrained Fay-Herriot model for the last one deletion benchmarking (see Section 1).

Theorem 2

Let $\theta_i \overset{\text{ind}}{\sim} \text{Normal}\left(\mathbf{u}_i'\boldsymbol{\beta}, \delta^2\right)$, $i = 1, \ldots, \ell$. Then, under the constraint, $\sum_{i=1}^{\ell} \theta_i = a$, where a is constant, letting $\boldsymbol{\theta}_{(\ell)}$ be the vector of all the θ_i except the last one, the joint density of θ_i, $i = 1, \ldots, \ell$, is

$$\boldsymbol{\theta}_{(\ell)} \sim \text{Normal}\left\{\left(I - \frac{1}{\ell} J\right)\mathbf{c},\ \delta^2\left(I - \frac{1}{\ell} J\right)\right\},\ \theta_\ell = a - \sum_{i=1}^{\ell-1} \theta_i, \tag{3.1}$$

$\mathbf{c}' = a\mathbf{j}' + \left((\mathbf{u}_1 - \mathbf{u}_\ell)'\boldsymbol{\beta}, \ldots, (\mathbf{u}_{\ell-1} - \mathbf{u}_\ell)'\boldsymbol{\beta}\right)$, J and $\mathbf{j}$ are respectively a $(\ell - 1) \times (\ell - 1)$ matrix and a $(\ell - 1)$ vector of ones.

Proof of Theorem 2

See Appendix C.

The proof of Theorem 2 uses the multivariate normal distribution, and it will be used to prove the more general theorem when the prior is adjusted to delete any area.

The constrained BFH model is

$$\hat{\theta}_i \,|\, \theta_i \overset{\text{ind}}{\sim} \text{Normal}\left(\theta_i, s_i^2\right),\ i = 1, \ldots, \ell, \tag{3.2}$$

$$\theta_i \,|\, \boldsymbol{\beta}, \sigma^2 \overset{\text{ind}}{\sim} \text{Normal}\left(\mathbf{x}_i'\boldsymbol{\beta}, \sigma^2\right),\ \sum_{i=1}^{\ell} \theta_i = a,$$

$$\pi\left(\boldsymbol{\beta}, \sigma^2\right).$$

The constraint on the prior on $\theta_1, \ldots, \theta_\ell$ essentially adjusts the joint prior density. However, to incorporate the constraint, we will use Theorem 2 to adjust the posterior density under the unconstrained model.

For $i = 1, \ldots, \ell$, let $\lambda_i = \frac{\sigma^2}{\sigma^2 + s_i^2}$ and $\hat{\hat{\theta}}_i = \lambda_i \hat{\theta}_i + (1 - \lambda_i)\, \mathbf{x}_i' \boldsymbol{\beta}$. Also, let $\hat{\sigma}_i^2 = \sigma^2 (1 - \lambda_i)$, $i = 1, \ldots, \ell$. Then, under the unconstrained model the joint posterior density is

$$\pi_{nb}\left(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2 \,|\, \hat{\boldsymbol{\theta}}\right) \propto \pi(\boldsymbol{\beta}, \sigma^2) \prod_{i=1}^{\ell} \left[\text{Normal}_{\hat{\theta}_i}\left(\mathbf{x}_i' \boldsymbol{\beta}, \frac{\sigma^2}{\lambda_i}\right) \text{Normal}_{\theta_i}\left(\hat{\hat{\theta}}_i, \hat{\sigma}_i^2\right) \right].$$

We now extend the result of Theorem 2, to reflect our interest in the constrained BFH model. Theorem 3, below, is used to construct the random deletion benchmarking method.

Theorem 3

Using a general notation, let $y_i \overset{\text{ind}}{\sim} \text{Normal}(\mu_i, \sigma_i^2)$, $i = 1, \ldots, \ell$. Let $\mathbf{y}_{(i)}$ denote the vector of all the y_i except the i^{th} one. Also, let $v_i = \sigma_i^2 \Big/ \sqrt{\sum_{j=1}^{\ell} \sigma_j^2}$, $i = 1, \ldots, \ell$, and $v_i^* = \sigma_i^2 \Big/ \sum_{j=1}^{\ell} \sigma_j^2$, $i = 1, \ldots, \ell$. Under the constraint $\sum_{i=1}^{\ell} y_i = a$,

$$\mathbf{y}_{(i)} \sim \text{Normal}\left\{ \mathbf{u}_{(i)} - \left(\sum_{j=1}^{\ell} u_j - a \right) \mathbf{v}_{(i)}^*,\ \text{diagonal}\left(\boldsymbol{\sigma}_{(i)}^2\right) - \mathbf{v}_{(i)} \mathbf{v}_{(i)}' \right\} \tag{3.3}$$

with $y_i = a - \sum_{j \neq i} y_j$. Then,

$$p(\mathbf{y},\, z = r \,|\, \phi = 0) = \delta_{y_r}\left(a - \sum_{j \neq r} y_j \right) \text{Normal}_{\mathbf{y}_{(r)}} \left\{ \boldsymbol{\mu}_{(r)} - \left(\sum_{j=1}^{\ell} \mu_j - a \right) \mathbf{v}_{(r)}^*,\ \text{diagonal}\left(\boldsymbol{\sigma}_{(r)}^2\right) - \mathbf{v}_{(r)} \mathbf{v}_{(r)}' \right\}, \tag{3.4}$$

for $r = 1, \ldots, \ell$ and $\psi = a - \sum_{i=1}^{\ell} y_i$.

Proof of Theorem 3

The proof of Theorem 3 is similar to Theorem 2. See Appendix C.

In what follows, one of the ℓ area parameters will be deleted randomly (i.e., with probability $1/\ell$). Let $z = 1, \ldots, \ell$ represent the county that is deleted. That is, $P(z = r) = 1/\ell$, $r = 1, \ldots, \ell$. Then, under the constrained BFH model, using Theorem 3, the joint posterior density is

$$\begin{aligned} \pi_b\left(\boldsymbol{\theta}, z = r, \boldsymbol{\beta}, \sigma^2 \,|\, \hat{\boldsymbol{\theta}}\right) \ \propto\ & \pi(\boldsymbol{\beta}, \sigma^2) \prod_{i=1}^{\ell} \left[\text{Normal}_{\hat{\theta}_i}\left(\mathbf{x}_i' \boldsymbol{\beta}, \frac{\sigma^2}{\lambda_i}\right) \right] \times \delta_{\theta_r}\left(a - \sum_{j \neq r} \theta_j \right) \\ & \times \text{Normal}_{\boldsymbol{\theta}_{(r)}} \left\{ \hat{\hat{\boldsymbol{\theta}}}_{(r)} - \left(\sum_{j=1}^{\ell} \hat{\hat{\theta}}_j - a \right) \mathbf{v}_{(r)}^*,\ \text{diagonal}\left(\hat{\boldsymbol{\sigma}}_{(r)}^2\right) - \mathbf{v}_{(r)} \mathbf{v}_{(r)}' \right\}, \end{aligned} \tag{3.5}$$

where $r = 1, \ldots, \ell$, and for $i = 1, \ldots, \ell$, $v_i^* = \hat{\sigma}_i^2 \Big/ \sum_{j=1}^{\ell} \hat{\sigma}_j^2$ and $v_i = \hat{\sigma}_i^2 \Big/ \sqrt{\sum_{j=1}^{\ell} \hat{\sigma}_j^2}$.

3.2 Sampling the joint posterior density

Unlike the BFH model, the constrained model in (3.2) cannot be fit using random draws; we use a Gibbs sampler. The joint conditional posterior density (cpd) of $(\boldsymbol{\theta}, z)$ is

$$\pi\left(\boldsymbol{\theta},\, z = r \,|\, \boldsymbol{\beta},\, \sigma^2,\, \hat{\boldsymbol{\theta}}\right) \;\propto\; \delta_{\theta_r}\left(a - \sum_{j \neq r} \theta_j\right)$$
$$\times\; \text{Normal}_{\boldsymbol{\theta}_{(r)}}\left\{\hat{\hat{\boldsymbol{\theta}}}_{(r)} - \left(\sum_{j=1}^{\ell} \hat{\hat{\theta}}_j - a\right)\mathbf{v}^*_{(r)},\, \text{diagonal}\left(\hat{\boldsymbol{\sigma}}^2_{(r)}\right) - \mathbf{v}_{(r)}\mathbf{v}'_{(r)}\right\}, \qquad (3.6)$$

where $r = 1, \ldots, \ell$, $\boldsymbol{\theta}_{(r)}$ denotes the vector of θ_i with the r^{th} component deleted, and $\mathbf{v}$ and $\mathbf{v}^*$ are defined in Theorem 3. Then, the joint conditional posterior density of $(\boldsymbol{\beta},\, \sigma^2)$ is

$$\pi\left(\boldsymbol{\beta},\, \sigma^2 \,|\, \boldsymbol{\theta},\, \hat{\boldsymbol{\theta}},\, z = r\right) \;\propto\; \pi(\boldsymbol{\beta},\, \sigma^2) \prod_{i=1}^{\ell} \text{Normal}_{\hat{\theta}_i}\left(\mathbf{x}'_i\boldsymbol{\beta},\, \frac{\sigma^2}{\lambda_i}\right)$$
$$\times\; \text{Normal}_{\boldsymbol{\theta}_{(r)}}\left\{\hat{\hat{\boldsymbol{\theta}}}_{(r)} - \left(\sum_{j=1}^{\ell} \hat{\hat{\theta}}_j - a\right)\mathbf{v}^*_{(r)},\, \text{diagonal}\left(\hat{\boldsymbol{\sigma}}^2_{(r)}\right) - \mathbf{v}_{(r)}\mathbf{v}'_{(r)}\right\}. \qquad (3.7)$$

It is straight forward to sample the cpd's of $\boldsymbol{\theta}$ and z. However, it is not so simple to sample the cpd's of $\boldsymbol{\beta}$ and σ^2 that we will next discuss.

First, to obtain the cpd's of $\boldsymbol{\beta}$ and σ^2, we define

$$\mathbf{g}_1 = \frac{1}{\sigma^2}\sum_{i=1}^{\ell} \lambda_i \hat{\theta}_i \mathbf{x}_i \quad \text{and} \quad A_1 = \frac{1}{\sigma^2}\sum_{i=1}^{\ell} \lambda_i \mathbf{x}_i \mathbf{x}'_i.$$

Second, for $i = 1, \ldots, \ell$, let $d_i = \lambda_i\hat{\theta}_i - \left(\sum_{j=1}^{\ell} \lambda_j \hat{\theta}_j - a\right) v_i^*$ and $\mathbf{x}_i = (1 - \lambda_i)\,\mathbf{x}_i - v_i^* \sum_{j=1}^{\ell} (1 - \lambda_j)\,\mathbf{x}_j$. Let $\mathbf{d}_{(r)}$ and $\tilde{X}_{(r)}$ respectively denote the vector with entries d_i excluding the r^{th} component and the matrix with columns $\mathbf{x}_i$ excluding $\mathbf{x}_r$. Now, let

$$\mathbf{g}_2 = \left(\boldsymbol{\theta}_{(r)} - \mathbf{d}_{(r)}\right)' \Sigma_{(r)} \tilde{X}'_{(r)} \quad \text{and} \quad A_2 = \tilde{X}_{(r)} \Sigma_{(r)} \tilde{X}'_{(r)},$$

where $\Sigma_{(r)}$ is the covariance matrix without the r^{th} row and column. Third, with a multivariate normal prior on $\boldsymbol{\beta}$ of the form $\boldsymbol{\beta} \sim \text{Normal}(\boldsymbol{\beta}_0,\, \Sigma_0)$, where $\boldsymbol{\beta}_0$ and Σ_0 are specified, we let

$$\mathbf{g}_0 = \boldsymbol{\beta}_0 A_0 \quad \text{and} \quad A_0 = \Sigma_0^{-1};$$

thereby offering some protection against posterior impropriety. It follows that

$$\boldsymbol{\beta} \,|\, \boldsymbol{\theta},\, z = r,\, \sigma^2,\, \hat{\boldsymbol{\theta}} \sim \text{Normal}\left\{\left(\sum_{s=0}^{2} A_s\right)^{-1} \sum_{s=0}^{2} A_s \mathbf{g}_s,\, \left(\sum_{s=0}^{2} A_s\right)^{-1}\right\}.$$

We can eliminate the prior of $\boldsymbol{\beta}$ by letting $\Sigma_0 \to \infty$ (i.e., noninformative prior) to get $\pi(\boldsymbol{\beta}) = 1$ as in the BFH model.

Finally, we consider the cpd of σ^2. Let

$$U_i = \lambda_i\hat{\theta}_i + (1 - \lambda_i)\,\mathbf{x}'_i\boldsymbol{\beta} - \left\{\sum_{j=1}^{\ell}\left\{\lambda_j\hat{\theta}_j + (1 - \lambda_j)\,\mathbf{x}'_j\boldsymbol{\beta}\right\} - a\right\} v_i^*,\; i = 1, \ldots, \ell,$$

and

$$\Sigma = \sigma^2\left\{\text{diagonal}\,(1 - \lambda_1, \ldots, 1 - \lambda_\ell) - \left[(1 - \lambda_i)(1 - \lambda_{i'})\right] \Big/ \sum_{j=1}^{\ell}(1 - \lambda_j)\right\}.$$

Then, the cpd of σ^2 is

$$\pi\left(\sigma^2 \,|\, \boldsymbol{\theta},\, z = r,\, \boldsymbol{\beta},\, \hat{\boldsymbol{\theta}}\right) \propto \pi(\sigma^2)\left[\prod_{i=1}^{\ell} \text{Normal}_{\hat{\theta}_i}\left\{\mathbf{x}_i'\boldsymbol{\beta},\, \sigma^2 + s_i^2\right\}\right] \text{Normal}_{\boldsymbol{\theta}_{(r)}}\left\{\mathbf{U}_{(r)},\, \Sigma_{(r)}\right\},$$

where $\mathbf{U}_{(r)}$ denotes the vector of the U_i excluding the r^{th} component and $\pi(\sigma^2)$ is a prior on σ^2. As in the BFH model (see Section 2), we assign the prior density $\pi(\sigma^2) = 1/(1+\sigma^2)^2$, $\sigma^2 > 0$ to σ^2. Because the baseline BFH posterior density is proper, the constraint BFH posterior density will also be proper.

4 Empirical studies

The purposes of these empirical studies are twofold. First, it is demonstrated that the BFH model can be fit as stated in Section 2 and the deleting the last one benchmarking and random benchmarking methods are performed. Second, the benchmarking methods are compared in a simulation study that uses a well-used dataset in the small area literature.

In the data generation process, we use the data on corn and soybean acres in Battese, Harter and Fuller (1988), available for 12 counties (areas) in Iowa. The resulting county-level corn and soybean acreages are constructed using a number of segments sampled from the population (known number of segments). Landsat satellite data on the number of pixels of corn and soybean in the sampled segments (i.e., two covariates) are also available. The finite population means of the number of pixels classified as corn and soybean for each county are also reported. Starting with this dataset, we construct new datasets with any number of areas.

The data generation process has two steps. In the first step, the unit-level model $y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + e_{ij}$, $i = 1, \ldots, \ell,\ j = 1, \ldots, n_i$, where $e_{ij} \overset{\text{iid}}{\sim} (0,\, \sigma^2)$, is fit to the data available for the $\ell = 12$ counties in Iowa. The area sample sizes are $n_1 = n_2 = n_3 = 1$, $n_4 = 2$, $n_5 = n_6 = n_7 = n_8 = 3$, $n_9 = 4$, $n_{10} = n_{11} = 5$, and $n_{12} = 6$. Using least squares, we estimate $\boldsymbol{\beta}$ and σ^2 by $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$, respectively. For the areas with sample size greater than one, we set s_i^2 equal to the estimated variance of the sample mean $\bar{y}_i$ $\left(\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i\right)$ and we let S^2 be their geometric mean. For the areas with sample size equal to one, we set s_i^2 equal to S^2. The vector of covariates $\bar{\mathbf{X}}_i$ has three elements, the integer one (for the intercept), followed by the population means of pixels classified as corn and soybean.

In the second step, the data generation process for any desired number ℓ of small areas is illustrated. The covariates $\mathbf{x}_i$, $i = 1, \ldots, \ell$, are sampled with replacement from $\bar{\mathbf{X}}_i$, $i = 1, \ldots, 12$. Then, the area-level means are drawn using

$$\theta_i \overset{\text{ind}}{\sim} \text{Normal}\left(\mathbf{x}_i'\hat{\boldsymbol{\beta}},\, \hat{\sigma}^2\right),\ i = 1, \ldots, \ell,$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are the least squares estimates defined above. The sample variances s_i^2 are generated in two steps. First, the sample sizes are drawn from a uniform distribution, $n_i \overset{\text{iid}}{\sim} \text{Uniform}\,(5,\, 25)$, $i = 1, \ldots, \ell$. Second, let $s_i^2 = S^2 V_i/(n_i - 1)$, where $V_i \overset{\text{ind}}{\sim} \chi^2_{n_i - 1}$ and S^2 defined above. Finally, the small area survey estimates are drawn using $\hat{\theta}_i \overset{\text{ind}}{\sim} \text{Normal}\,(\theta_i,\, s_i^2)$, $i = 1, \ldots, \ell$. The benchmarking target is set equal to the sum of the $\hat{\theta}_i$ and variants of this value, $\sum_{i=1}^{\ell} \hat{\theta}_i$ scaled up or down by 50%. In NASS's practice,

for crop county estimates, this target is an already set state value. To evaluate the benchmarking methods in extreme cases, we consider additional simulation scenarios, where an area sample size is set to 2 or 50, or where the factor S^2 is multiplied by ten.

In what follows, we report empirical results mostly for a simulation scenario using 12 areas. Examples using larger number of areas are briefly discussed. For example, Iowa has 99 counties, and one of NASS's interests is in benchmarking county estimates for planted acres, harvested acres and production (bushels) to the predefined state-level total. For such small numbers of areas, no adjustment is needed to the benchmarking procedures, deleting the last one or random deletion, introduced in the previous sections. However, the computation may be intolerable for an extremely large number of areas (say, one million), and some adjustments would be needed to the current procedures.

It is pertinent to discuss the computations for the simulation scenario with 12 areas. For posterior inference under the BFH model, we have used 1,000 random draws, and this runs in just a few seconds. On the other hand, it is more difficult to run a Gibbs sampler for deleting one at a time or random deletion benchmarking. However, we have provided an efficient Gibbs sampler as follows. We used a long run of 20,000 iterations, with a "burn in" of the first 10,000 iterations, choosing every tenth iterate thereafter. This was obtained by trial and error that is gauged by the autocorrelations, the Geweke test for stationarity and the effective sample sizes. For the 1,000 selected iterations, the autocorrelations are all negligible. For random deletion benchmarking, the p-values of the Geweke test for the three regression coefficients and δ^2 are, respectively, 0.651, 0.087, 0.828 and 0.699 (i.e., stationarity is not rejected), and the effective sample sizes are all 1,000. Also, the trace plots show no evidence of nonstationarity. Therefore, the Gibbs sampler is efficient, taking a few seconds despite the large number of runs.

The performance of benchmarking methods is assessed using a set of metrics that include posterior means (PM) and posterior standard deviations (PSD), and when it is convenient, posterior coefficients of variation (PCV), numerical standard errors (NSE) of the estimates and 95% highest posterior density intervals (95% HPD). Numerical results are presented in Tables 4.1-4.8.

A summarized version of the basic results is presented in Table 4.1, and serves for comparison of the average, standard error and coefficient of variation of the observed data with the PMs, PSDs, PCVs from the BFH model, benchmarking (deleting the last one, LO) model and random benchmarking (RD) model. The results in Table 4.1 apply to two simulation scenarios, where $S^2 = 163,$ small variation in the observed data, and where $S^2 = 1{,}630,$ relatively larger variation in the observed data. When $S^2 = 163,$ there are very little differences between the observed data and the posterior quantities from the BFH, LO and RD models. Given the small coefficients of variation for the survey estimates, it is difficult for any model to further reduce variability. Hence, the PCVs are comparable to the CVs of the survey estimates. On the other hand, three interesting points can be made for the scenario where $S^2 = 1{,}630.$ First, the PMs under the BFH model can be very different from those of LO and RD models and these latter two PMs are very close. Second, the PSDs are much smaller than the standard errors of the observed data; there are substantial gains in precision under the BFH model. However, the PSDs are about four to five times smaller than those for the observed data and the PSDS under the LO and RD model are about twice those of the BFH model. The

PCVs follow the same pattern. Third, LO and RD are very close in all three measures (PMs, PSDs, PCVs) with RD model having just slightly smaller PSDs. As expected, there is small difference between the LO model and the RD model. But one must also observe that benchmarking the BFH model is important because we can get answers that are different from the BFH model at least in terms of posterior standard deviations and coefficients of variation. Benchmarking is a jittering procedure, which helps to protect the model from misspecification, and therefore it must lead to increased variability in the small area estimates.

Table 4.1
Comparison of BFH model with no benchmarking, deleting the last one benchmarking and random benchmarking via posterior mean (PM), posterior standard deviation (PSD) and posterior coefficient of variation (PCV) for two values of S^2

		PM				PSD				PCV			
	A	OB	BFH	LO	RD	OB	BFH	LO	RD	OB	BFH	LO	RD
a. $S^2 = 163$; $a = 1{,}435$	1	135.6	134.0	133.8	133.5	6.03	5.62	5.47	5.41	0.044	0.042	0.041	0.041
	2	102.0	103.5	103.1	103.0	7.10	6.50	6.11	5.82	0.070	0.063	0.059	0.057
	3	117.7	121.0	120.7	120.5	7.31	6.72	6.55	6.25	0.062	0.056	0.054	0.052
	4	77.0	81.5	81.4	81.0	5.88	6.00	5.46	5.53	0.076	0.074	0.067	0.068
	5	126.9	127.8	127.5	127.5	5.63	5.25	5.25	5.06	0.044	0.041	0.041	0.040
	6	113.1	113.4	112.9	113.1	8.06	7.15	6.82	6.74	0.071	0.063	0.060	0.060
	7	137.2	133.7	133.5	133.9	6.74	6.38	5.93	6.02	0.049	0.048	0.044	0.045
	8	124.8	124.7	124.7	124.7	4.03	3.91	3.83	3.76	0.032	0.031	0.031	0.030
	9	118.3	116.5	115.8	116.6	7.54	6.79	6.29	6.65	0.064	0.058	0.054	0.057
	10	156.5	153.4	153.3	153.3	4.37	4.45	4.12	4.18	0.028	0.029	0.027	0.027
	11	109.5	110.3	110.3	110.2	4.88	4.64	4.70	4.70	0.045	0.042	0.043	0.043
	12	116.3	118.1	117.9	117.7	7.23	6.62	6.26	6.00	0.062	0.056	0.053	0.051
b. $S^2 = 1{,}630$; $a = 1{,}482$	1	129.1	129.8	127.2	126.5	19.07	4.64	10.71	10.45	0.148	0.036	0.084	0.083
	2	117.3	126.3	122.1	122.1	22.46	5.08	12.73	12.51	0.191	0.040	0.104	0.102
	3	120.0	145.5	137.3	136.9	23.11	5.93	12.91	12.68	0.193	0.041	0.094	0.093
	4	68.8	107.3	94.0	93.6	18.60	7.47	12.04	11.86	0.270	0.070	0.128	0.127
	5	142.4	146.4	142.3	142.2	17.80	4.52	11.98	11.15	0.125	0.031	0.084	0.078
	6	108.8	120.2	115.2	115.4	25.49	5.43	11.75	11.66	0.234	0.045	0.102	0.101
	7	136.8	116.2	118.2	119.0	21.31	5.37	11.32	11.90	0.156	0.046	0.096	0.100
	8	124.5	132.5	127.3	127.3	12.76	4.39	9.00	8.91	0.102	0.033	0.071	0.070
	9	144.2	127.5	128.0	129.5	23.86	5.33	12.74	14.00	0.165	0.042	0.100	0.108
	10	172.9	129.2	145.5	145.3	13.81	9.23	10.28	10.37	0.080	0.071	0.071	0.071
	11	109.1	114.7	110.6	110.2	15.42	4.31	10.53	10.43	0.141	0.038	0.095	0.095
	12	108.4	120.3	114.6	114.2	22.87	5.10	12.42	12.01	0.211	0.042	0.108	0.105

Note: OB: observed data; BFH: Bayesian Fay-Herriot model; LO: benchmarking (deleting the last one) model; RD: random benchmarking model; a is the target. For OB, the direct estimate, standard error and coefficient of variation are presented under PM, PSD and PCV, respectively. Under the DGSM benchmarking procedure, at $S^2 = 163$, the benchmarking values are 133.7, 103.3, 120.8, 81.3, 127.6, 113.2, 133.4, 124.5, 116.3, 153.1, 110.1, 117.9, and at $S^2 = 1{,}630$, the benchmarking values are 126.9, 123.5, 142.3, 105.0, 143.2, 117.5, 113.6, 129.5, 124.7, 126.4, 112.1, 117.6.

Under the basic simulation scenario, we compare the deletion benchmarking methods to one of the methods in DGSM that provides benchmarked posterior estimates without deletion. To match the notation in DGSM, the benchmarking equation must be rewritten as

$$\sum_{i=1}^{\ell} \omega_i \theta_i = \frac{a}{\ell} = t,$$

where $\omega_i = 1/\ell$, $\sum_{i=1}^{\ell} \omega_i = 1$. Let $\hat{\theta}_i^{(B)}$ denote the posterior means from the BFH model. Now, define $\bar{\hat{\theta}}_B = \sum_{i=1}^{\ell} \omega_i \hat{\theta}_i^{(B)}$,

$$\phi_i = \frac{\omega_i}{\hat{\theta}_i^{(B)}},\ r_i = \frac{\omega_i}{\phi_i},\ i = 1, \ldots, \ell,$$

and $S^* = \sum_{i=1}^{\ell} \omega_i^2 / \phi_i$. Note that among the several specifications in DGSM, we have selected ϕ_i at random (no preference). Then, the benchmarked Bayes estimators of DGSM are

$$\hat{\theta}_i^{(BM)} = \hat{\theta}_i^{(B)} + \left(t - \bar{\hat{\theta}}_B\right) r_i / S^*,\ i = 1, \ldots, \ell.$$

Empirical results using the estimator $\hat{\theta}_i^{(BM)}$ are presented in the note to Table 4.1. The largest difference between the benchmarked estimates under different benchmarking methods is for area 10 (OB: 172.9; BFH: 129.2; LO: 145.5; RD: 145.3; DGSM: 126.4). In general, the PMs from LO and RD are closer to OB (observed data). Otherwise, these estimates compare reasonably well with the LO benchmarking and RD deletion although there are some small differences; DGSM does not provide posterior standard deviations and credible intervals.

More detailed results for $S^2 = 163$, are presented in Tables 4.2-4.8 and in Figures 4.1-4.4. Our interest is mainly to compare deletion of a single area (e.g., LO) and RD.

Using the results in Table 4.2, we conclude that the PMs from the BFH model (without benchmarking) are slightly different from the direct estimates, and as expected, larger than the smaller direct estimates and smaller than the larger ones. Except for two areas, as expected, the PSDs are smaller than the direct standard deviations. For example, the smallest direct estimate (76.997) has the largest shrinkage with a larger standard deviation (5.881 vs. 5.995); the results are consistent with the standard shrinkage that occurs in small area estimation. We note that the PCVs are all small and the NSEs are reasonably small, too.

Table 4.2
Comparison of the direct estimator with posterior inference from the Bayesian Fay-Herriot model for the area parameters

Area	n	$\hat{\theta}$	s	PM	PSD	PCV	NSE	95% HPD
1	5	135.575	6.031	133.985	5.617	0.042	0.057	(123.422, 145.402)
2	7	101.980	7.101	103.461	6.498	0.063	0.065	(90.598, 116.134)
3	24	117.655	7.309	121.006	6.716	0.056	0.066	(107.730, 134.124)
4	23	76.997	5.881	81.473	5.995	0.074	0.058	(69.046, 92.578)
5	21	126.917	5.629	127.832	5.248	0.041	0.052	(117.850, 138.406)
6	9	113.132	8.061	113.393	7.147	0.063	0.068	(99.441, 127.451)
7	5	137.236	6.739	133.661	6.378	0.048	0.064	(121.771, 146.662)
8	20	124.839	4.034	124.732	3.906	0.031	0.039	(117.233, 132.309)
9	16	118.306	7.544	116.479	6.785	0.058	0.071	(103.225, 130.003)
10	9	156.503	4.368	153.355	4.449	0.029	0.045	(144.785, 162.031)
11	23	109.546	4.877	110.348	4.637	0.042	0.047	(101.179, 119.294)
12	9	116.314	7.232	118.098	6.623	0.056	0.068	(105.135, 131.186)

Note: n is the area sample size, $\hat{\theta}$ is the direct estimator and s its standard error. PM is the posterior mean, PSD is the posterior standard deviation and HPD is highest posterior density interval. NSE is the numerical standard errors of the posterior means. The benchmarking value is 1,435 and the sum of the posterior mean is 1,437.823 (not benchmarked).

The estimates from the BFH model with deleting the last area and with random deletion under a uniform prior (equal weights) are presented in Tables 4.3 and 4.4. The posterior weights barely differ from 0.083

with the largest one (0.097) of the last area and smallest one (0.056) of the 8^{th} area. Both random deletion and deleting the last one provide improved precision, as the PSDs of the benchmarked estimates are all smaller than the observed standard errors, for both benchmarking methods. The NSEs are larger than for no benchmarking, but this barely matters as these are errors of the PMs (the characteristic of the PM has three digits).

Table 4.3
Comparison of the direct estimator with posterior inference from the Bayesian Fay-Herriot model for the area parameters under random deletion benchmarking

Area	n	$\hat{\theta}$	s	PM	PSD	PCV	NSE	95% HPD
1	5	135.575	6.031	133.516	5.431	0.041	0.171	(123.414, 143.541)
2	7	101.980	7.101	102.903	5.793	0.056	0.199	(92.378, 114.250)
3	24	117.655	7.309	120.671	6.237	0.052	0.194	(107.744, 132.190)
4	23	76.997	5.881	81.170	5.597	0.069	0.202	(69.781, 91.177)
5	21	126.917	5.629	127.652	5.036	0.039	0.170	(118.293, 137.228)
6	9	113.132	8.061	112.805	6.707	0.059	0.223	(100.926, 126.074)
7	5	137.236	6.739	133.908	6.007	0.045	0.177	(122.135, 145.344)
8	20	124.839	4.034	124.703	3.757	0.030	0.120	(117.962, 132.304)
9	16	118.306	7.544	116.451	6.650	0.057	0.249	(103.400, 129.316)
10	9	156.503	4.368	153.222	4.216	0.028	0.134	(144.392, 160.854)
11	23	109.546	4.877	110.221	4.694	0.043	0.150	(101.038, 119.570)
12	9	116.314	7.232	117.780	5.997	0.051	0.208	(104.619, 128.158)

Note: n is the area sample size, $\hat{\theta}$ is the direct estimator and s its standard error. PM is the posterior mean, PSD is the posterior standard deviation and HPD is highest posterior density interval. NSE is the numerical standard errors of the posterior means. The benchmarking value is 1,435. Under a uniform prior (equal weights) the posterior probabilities that the areas 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 are deleted are respectively 0.090, 0.084, 0.095, 0.077, 0.066, 0.093, 0.097, 0.056, 0.098, 0.068, 0.079, 0.097.

Table 4.4
Comparison of the direct estimator with posterior inference from the Bayesian Fay-Herriot model for the area parameters under deleting the last area

Area	n	$\hat{\theta}$	s	PM	PSD	PCV	NSE	95% HPD
1	5	135.575	6.031	133.772	5.519	0.041	0.151	(122.213, 143.991)
2	7	101.980	7.101	103.026	6.319	0.061	0.171	(89.424, 113.857)
3	24	117.655	7.309	120.470	6.458	0.054	0.209	(108.783, 134.261)
4	23	76.997	5.881	81.391	5.906	0.073	0.171	(69.636, 92.634)
5	21	126.917	5.629	127.883	5.158	0.040	0.142	(117.282, 137.305)
6	9	113.132	8.061	112.895	6.270	0.056	0.216	(100.664, 124.320)
7	5	137.236	6.739	133.298	5.948	0.045	0.178	(121.831, 144.727)
8	20	124.839	4.034	124.664	3.810	0.031	0.124	(117.321, 131.941)
9	16	118.306	7.544	116.542	6.531	0.056	0.203	(104.238, 129.622)
10	9	156.503	4.368	153.229	4.353	0.028	0.132	(144.443, 161.593)
11	23	109.546	4.877	109.997	4.563	0.041	0.168	(101.428, 118.953)
12	9	116.314	7.232	117.835	6.344	0.054	0.215	(106.421, 131.483)

Note: n is the area sample size, $\hat{\theta}$ is the direct estimator and s its standard error. PM is the posterior mean, PSD is the posterior standard deviation and HPD is highest posterior density interval. NSE is the numerical standard errors of the posterior means. The benchmarking value is 1,435.

The three methods (BFH, RD, LO) are compared using the results in Table 4.5. The PMs are comparable, so that benchmarking (RD, LO) does not distort (shrink) the estimates much beyond the shrinkage under the BFH model. Also, the PSDs under LO and RD are almost always smaller than those under the BFH model. For eight of the twelve areas, RD has smaller PSDs than LO; in these areas, RD shows roughly 1% decrease in PSD over LO and roughly 4% over the PSDs from BFH.

To investigate how sensitive the PSDs are to different benchmarking targets, we present results using three choices of targets in Table 4.6. The PSDs change only slightly over different targets and are still better than the standard errors of the direct estimates.

As part of designing a complex set of simulations, we consider using unequal probabilities (weights) in the random deletion benchmarking, and present results in Table 4.7. Uniform weights (EW) are compared to weights inversely proportional (IW) to the sample sizes and to weights directly proportional (DW) to the samples sizes. Again, small differences are present among the three PMs and among the three PSDs. The PSDs are still smaller than those of the direct estimates.

Using the results in Table 4.8, we study how extreme sample sizes in the last county (to be deleted) affect posterior inference. For this, we set the sample size of the last county to be outside the simulation range (5-25), at 2 and 50. First, consider the case in which the sample size of the last county is 2. Consistent with previous findings, there are minor differences of the PMs over no benchmarking, deleting the last one and random deletion for all counties. The PSDs for LO and RD are smaller than those of BFH with nine of these PSDs for RD smaller than LO. However, for the last county, we observe relatively large posterior standard deviations (10.00, 8.771, 8.525), roughly 15% decrease in PSD of RD over no benchmarking. Next, consider the case in which the sample size of the last county is 50. The patterns are similar, except the PSDs for the last county are comparable to the others under BFH, LO and RD and again there is an approximately 10% decrease (6.282, 5.958, 5.702) in PSD of RD over no benchmarking. It appears that deliberately putting the county with the most extreme sample size (small or large) as the last county can affect the benchmarking procedure. In contrast, minor changes are observed when the areas with extreme sample size are not systematically deleted. When the sample size is 2, the new PMs and PSDs are the following, BFH: 124.307, 9.993; LO: 123.371, 9.000 RD: 123.540, 8.887. When the sample size is 50, the new PMs and PSDs are the following, BFH: 118.167, 6.284; LO: 117.802, 6.094; RD: 117.716, 5.948.

Table 4.5
A summary of the comparison of inference from the direct estimator, the Bayesian Fay-Herriot (BFH) model, random deletion (RD) benchmarking and deleting the last one (LO)

				BFH		RD		LO	
Area	n	$\hat{\theta}$	s	PM	PSD	PM	PSD	PM	PSD
1	5	135.575	6.031	133.985	5.617	133.516	5.431	133.772	5.519
2	7	101.980	7.101	103.461	6.498	102.903	5.793	103.026	6.319
3	24	117.655	7.309	121.006	6.716	120.671	6.237	120.470	6.458
4	23	76.997	5.881	81.473	5.995	81.170	5.597	81.391	5.906
5	21	126.917	5.629	127.832	5.248	127.652	5.036	127.883	5.158
6	9	113.132	8.061	113.393	7.147	112.805	6.707	112.895	6.270
7	5	137.236	6.739	133.661	6.378	133.908	6.007	133.298	5.948
8	20	124.839	4.034	124.732	3.906	124.703	3.757	124.664	3.810
9	16	118.306	7.544	116.479	6.785	116.451	6.650	116.542	6.531
10	9	156.503	4.368	153.355	4.449	153.222	4.216	153.229	4.353
11	23	109.546	4.877	110.348	4.637	110.221	4.694	109.997	4.563
12	9	116.314	7.232	118.098	6.623	117.780	5.997	117.835	6.344

Note: n is the area sample size, $\hat{\theta}$ is the direct estimator and s its standard error. PM is the posterior mean and PSD is the posterior standard deviation. The benchmarking value is 1,435. Under a uniform prior, the posterior probabilities that the areas 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 are deleted are respectively 0.090, 0.084, 0.095, 0.077, 0.066, 0.093, 0.097, 0.056, 0.098, 0.068, 0.079, 0.097.

Table 4.6
Comparison of posterior inference of the area parameters under random deletion benchmarking with different targets (a = 1,435)

				a		1.5a		0.5a	
Area	n	$\hat{\theta}$	s	PM	PSD	PM	PSD	PM	PSD
1	5	135.575	6.031	133.516	5.431	189.249	5.385	77.769	5.561
2	7	101.980	7.101	102.903	5.793	175.963	5.794	29.847	5.899
3	24	117.655	7.309	120.671	6.237	197.219	6.099	44.145	6.461
4	23	76.997	5.881	81.170	5.597	134.628	5.871	27.771	5.460
5	21	126.917	5.629	127.652	5.036	177.209	5.165	78.125	5.053
6	9	113.132	8.061	112.805	6.707	201.949	7.145	23.614	6.995
7	5	137.236	6.739	133.908	6.007	200.989	6.018	66.781	6.024
8	20	124.839	4.034	124.703	3.757	151.951	3.952	97.484	3.924
9	16	118.306	7.544	116.451	6.650	196.849	6.990	35.990	6.607
10	9	156.503	4.368	153.222	4.216	184.720	4.019	121.708	4.706
11	23	109.546	4.877	110.221	4.694	148.724	4.966	71.752	4.760
12	9	116.314	7.232	117.780	5.997	193.050	5.954	42.514	6.081

Note: n is the area sample size, $\hat{\theta}$ is the direct estimator and s its standard error. PM is the posterior mean and PSD is the posterior standard deviation. The benchmarking value is 1,435. Under a uniform prior, the posterior probabilities that the areas 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 are deleted are respectively 0.090, 0.084, 0.095, 0.077, 0.066, 0.093, 0.097, 0.056, 0.098, 0.068, 0.079, 0.097. When the benchmarking value is increased by 50%, these probabilities are 0.090, 0.084, 0.095, 0.079, 0.064, 0.093, 0.097, 0.056, 0.098, 0.068, 0.079, 0.097. When the benchmarking value is decreased by 50%, these probabilities are 0.090, 0.084, 0.095, 0.077, 0.066, 0.093, 0.097, 0.057, 0.097, 0.068, 0.079, 0.097.

Table 4.7
Comparison of posterior inference of the area parameters under random deletion benchmarking with equal weights (EW), weights inversely proportional sample sizes (IW) and weights directly proportional to sample sizes (DW)

				EW		IW		DW	
Area	n	$\hat{\theta}$	s	PM	PSD	PM	PSD	PM	PSD
1	5	135.575	6.031	133.516	5.431	133.508	5.518	133.436	5.404
2	7	101.980	7.101	102.903	5.793	103.042	5.737	103.049	5.809
3	24	117.655	7.309	120.671	6.237	120.529	6.176	120.634	6.247
4	23	76.997	5.881	81.170	5.597	81.167	5.571	81.111	5.567
5	21	126.917	5.629	127.652	5.036	127.669	5.079	127.541	5.055
6	9	113.132	8.061	112.805	6.707	112.762	6.704	113.074	6.716
7	5	137.236	6.739	133.908	6.007	133.965	5.968	133.798	6.027
8	20	124.839	4.034	124.703	3.757	124.829	3.734	124.719	3.757
9	16	118.306	7.544	116.451	6.650	116.300	6.707	116.502	6.640
10	9	156.503	4.368	153.222	4.216	153.238	4.198	153.204	4.220
11	23	109.546	4.877	110.221	4.694	110.190	4.697	110.208	4.690
12	9	116.314	7.232	117.780	5.997	117.802	6.010	117.726	5.989

Note: n is the area sample size, $\hat{\theta}$ is the direct estimator and s its standard error. PM is the posterior mean and PSD is the posterior standard deviation. The benchmarking value is 1,435. Under a uniform prior, the posterior probabilities that the areas 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 are deleted are respectively 0.090, 0.084, 0.095, 0.077, 0.066, 0.093, 0.097, 0.056, 0.098, 0.068, 0.079, 0.097. When the benchmarking is done using weights inversely proportional to sample sizes, these probabilities are 0.167, 0.127, 0.039, 0.037, 0.026, 0.105, 0.184, 0.030, 0.061, 0.078, 0.039, 0.107. When the benchmarking is done using weights directly proportional to sample sizes, these probabilities are 0.032, 0.048, 0.168, 0.124, 0.103, 0.061, 0.036, 0.083, 0.112, 0.044, 0.123, 0.066.

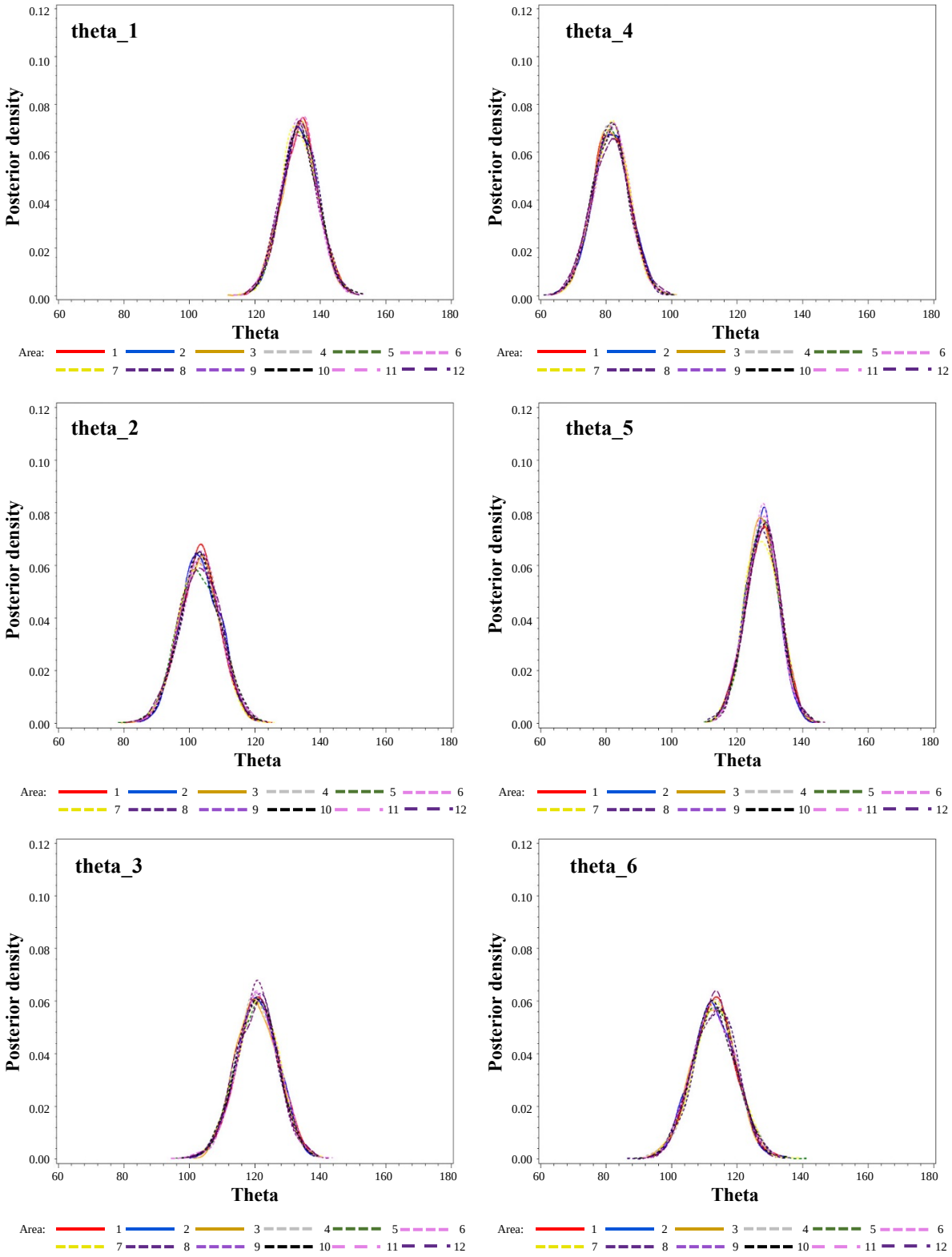
For comparison, different posterior densities are presented in Figures 4.1-4.4. In Figures 4.1 and 4.2, we present posterior densities of all twelve area parameters when each area, in turn, is deleted. We observe that the posterior densities are slightly different around the modes, but nothing remarkable. In Figures 4.3 and 4.4, we present posterior densities of all twelve area parameters under the FH model (unconstrained), random deletion benchmarking and deleting the last one. There are some differences among the three densities, but again these are not alarmingly different.

Finally, empirical results are presented for a simulation scenario with 99 areas, reflecting the 99 counties in Iowa. The data are generated as previously described, and the BFH model without benchmarking, with random deletion benchmarking, and with deleting the last one benchmarking is fit using 20,000 iterations for the Gibbs sampler. For each model fit, the first 10,000 iterations are used as a burn-in and every tenth iteration is kept thereafter. The BFH model fitting takes 15 seconds, while the deletion benchmarking models takes slightly less than three minutes each. For the random deletion benchmarking model parameters, the regression coefficients $\boldsymbol{\beta}$ and the variance σ^2, the p-values of the Geweke test are, respectively, 0.822, 0.128, 0.752 and 0.219, and the effective sample sizes are all 1,000 for the 1,000 selected iterations (i.e., an efficient Gibbs sampler). Note that the target is 12,162.93 and the sum of the PMs from the BFH model is 12,168.49, a difference of 5.56. In Figure 4.5, we present a plot of the coefficients of variation under random deletion benchmarking, deleting the last one benchmarking and BFH model versus the direct estimates by area. The differences among these models are not remarkable. Most of the points with direct CVs larger than about 0.04 fall below the 45^o straight line. However, some points (diamond) under the BFH model are above the 45^o line, four of them are noticeable, possibly shrinking too much. We conclude that it is sensible to perform the random deletion benchmarking.

Table 4.8
A summary of the comparison of inference from the direct estimator, the Bayesian Fay-Herriot (BFH) model, deleting the last one (LO) and random deletion (RD) benchmarking when the last county is extreme

					BFH		LO		RD	
	Area	n	$\hat{\theta}$	s	PM	PSD	PM	PSD	PM	PSD
a. The last county size is 2.	1	5	135.575	6.031	134.116	5.607	133.772	5.473	133.510	5.409
	2	7	101.980	7.101	103.205	6.482	102.818	6.118	102.745	5.837
	3	24	117.655	7.309	121.110	6.730	120.911	6.577	120.666	6.260
	4	23	76.997	5.881	81.586	6.021	81.741	5.544	81.196	5.631
	5	21	126.917	5.629	127.901	5.252	127.552	5.264	127.619	5.041
	6	9	113.132	8.061	113.454	7.147	112.889	6.818	113.074	6.815
	7	5	137.236	6.739	133.938	6.339	133.479	5.968	133.947	5.994
	8	20	124.839	4.034	124.753	3.906	124.699	3.824	124.738	3.735
	9	16	118.306	7.544	116.199	6.806	115.329	6.327	116.065	6.785
	10	9	156.503	4.368	153.419	4.434	153.148	4.174	153.240	4.213
	11	23	109.546	4.877	110.512	4.645	110.473	4.696	110.324	4.686
	12	2	121.881	12.75	124.243	10.00	123.755	8.771	123.444	8.525
b. The last county size is 50.	1	5	135.575	6.031	133.984	5.618	133.745	5.461	133.452	5.385
	2	7	101.980	7.101	103.462	6.499	103.136	6.086	103.044	5.780
	3	24	117.655	7.309	121.006	6.716	120.832	6.536	120.698	6.232
	4	23	76.997	5.881	81.473	5.995	81.596	5.512	81.162	5.728
	5	21	126.917	5.629	127.832	5.248	127.519	5.238	127.661	5.001
	6	9	113.132	8.061	113.393	7.146	112.929	6.777	112.899	6.675
	7	5	137.236	6.739	133.659	6.380	133.351	5.947	133.851	5.941
	8	20	124.839	4.034	124.732	3.906	124.713	3.821	124.726	3.825
	9	16	118.306	7.544	116.480	6.785	115.766	6.269	116.319	6.601
	10	9	156.503	4.368	153.355	4.449	153.225	4.173	153.306	4.230
	11	23	109.546	4.877	110.347	4.637	110.378	4.692	110.155	4.689
	12	50	116.538	6.791	118.117	6.282	118.035	5.958	117.952	5.702

Note: n is the area sample size, $\hat{\theta}$ is the direct estimator and s its standard error. PM is the posterior mean and PSD is the posterior standard deviation. When the sample size of the last county is 50 (2), the benchmarking value is 1,435 (1,441). The uniform prior is used in the random benchmarking.



Figures 4.1 Comparison of the posterior densities for θ_1 to θ_6 when each area is deleted at a time (e.g., the first area is deleted in the first panel etc.).

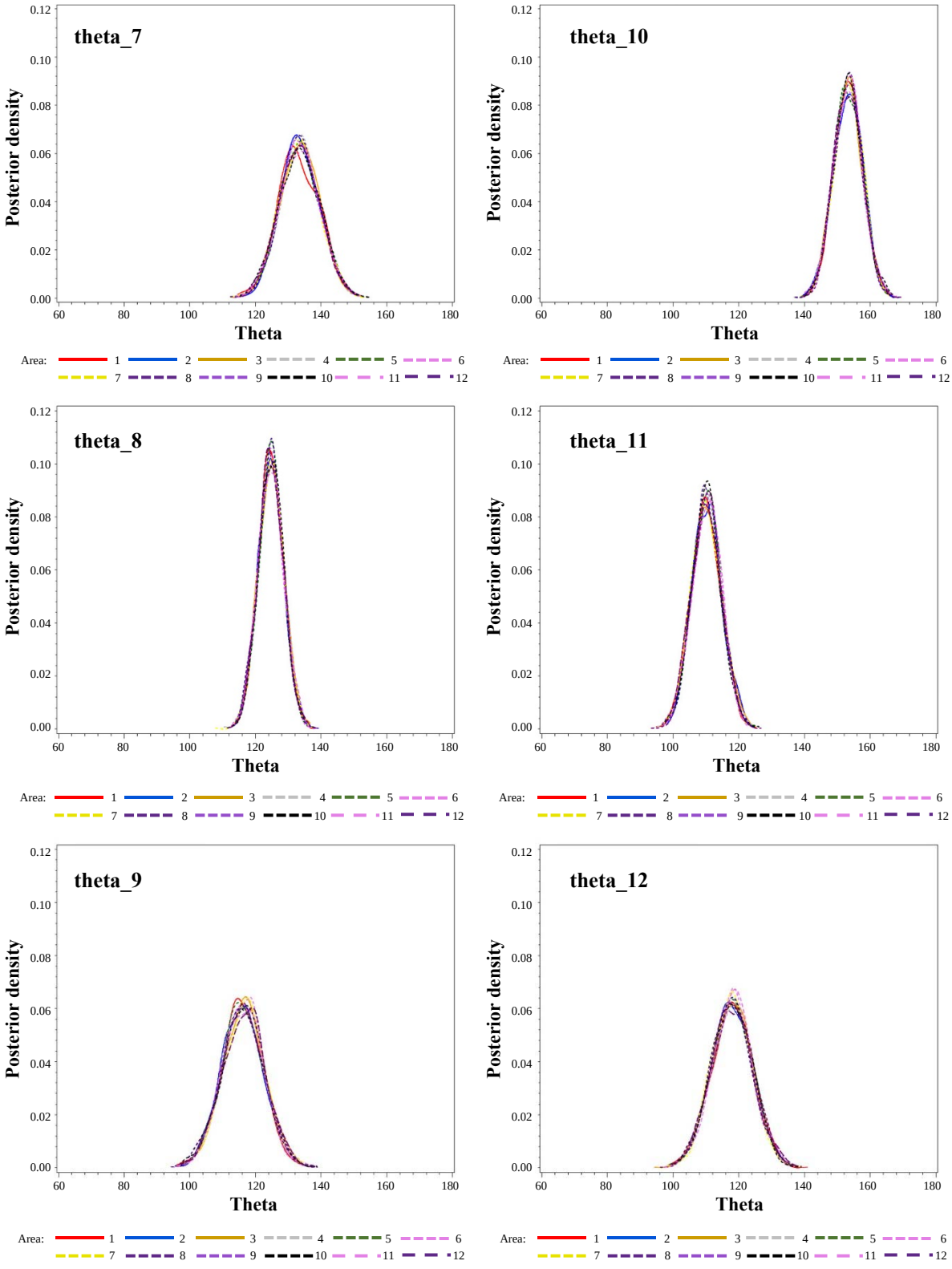


Figure 4.2 Comparison of the posterior densities for θ_7 to θ_{12} when each area is deleted at a time.

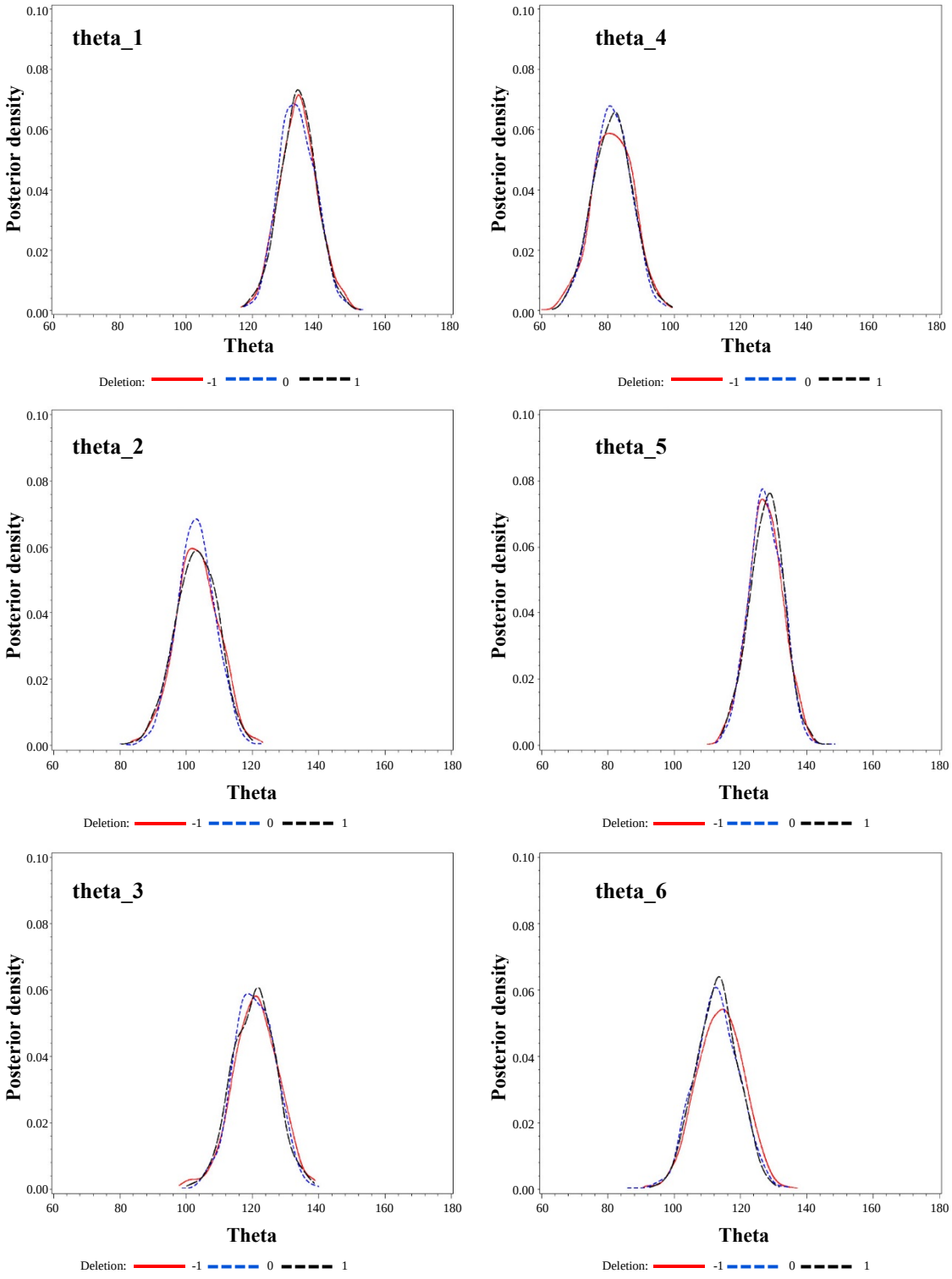


Figure 4.3 Comparison of the posterior densities for θ_1 to θ_6 under the Fay-Herriot model (-1), random deletion benchmarking (0) and area-12 deletion.

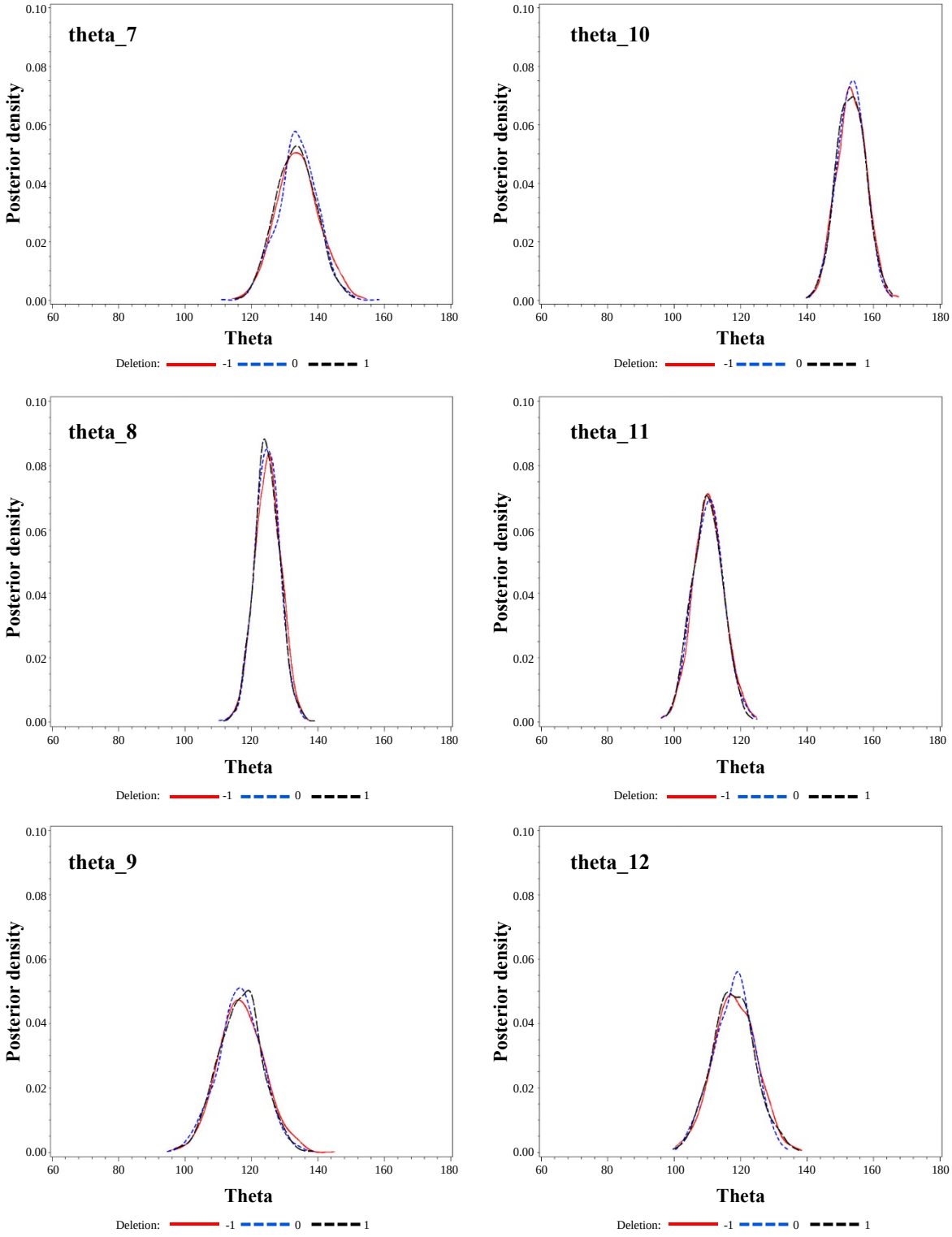


Figure 4.4 Comparison of the posterior densities of θ_7 to θ_{12} under the Fay-Herriot model (-1), random deletion benchmarking (0) and area-12 deletion.

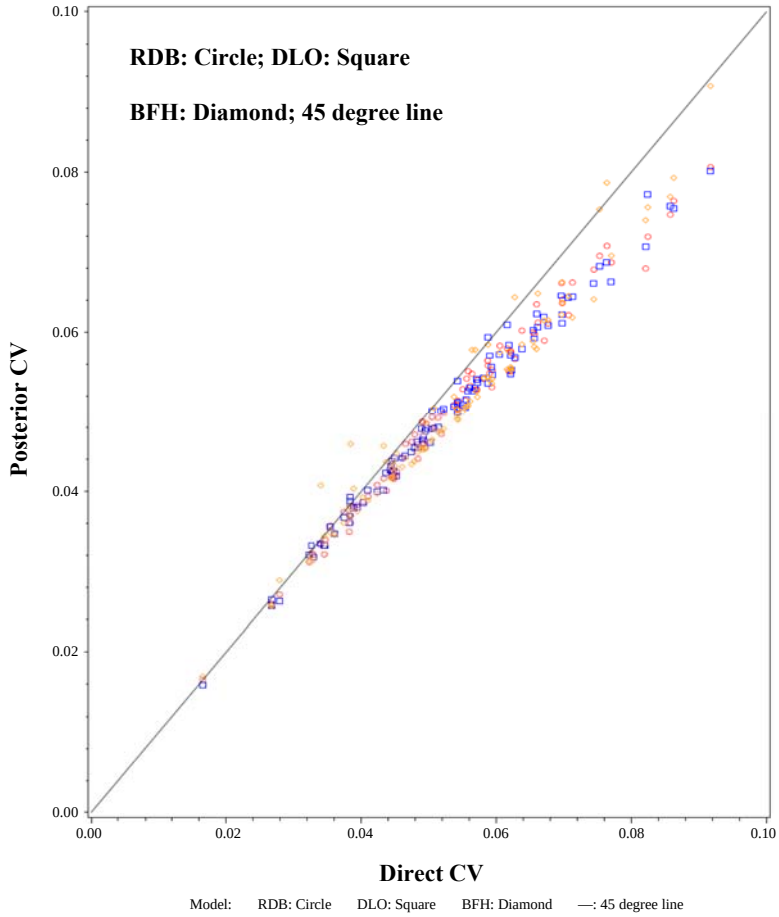


Figure 4.5 Plot of the coefficients of variation under the random deletion benchmarking, deleting the last one and the Bayesian Fay-Herriot model for 99 areas.

5 Concluding remarks

The Bayesian Fay-Herriot (BFH) model is discussed in detail. We show that the BFH can be fit using random samples rather than a Markov chain Monte Carlo sampler. Since random samples required no monitoring, this method is beneficial because there is little time at NASS between receiving the county-level survey summary data and presenting the final estimates. In support to the BFH model, we show that the posterior density under the BFH model is proper, providing a baseline for benchmarking. The effects of benchmarking are studied in a simulation study, comparing the BFH model without benchmarking to the BFH model with two benchmarking methods.

In this study, we assume that the benchmarking constraint is of the form $\sum_{i=1}^{\ell} \theta_i = a$. A straightforward generalization of the benchmarking methods may be developed for the constraint of the form $\sum_{i=1}^{\ell} w_i \theta_i = a$, where the w_i are weights. For example, this latter situation occurs for benchmarking yield, ratio of production and harvested acres.

Our major contribution is the extension of BFH model to accommodate benchmarking. Previous approaches delete the last area, giving rise to the question “Does it matter which area is deleted?”. In this paper, we develop and illustrate a method that gives each area a chance to be deleted. We show how to fit this extended BFH model using the Gibbs sampler. Because of the complexity of the joint posterior density, a sampling based method, without Markov chains, cannot be used. Using empirical studies, we show that the differences in the posterior means over no benchmarking, deleting the last county and random deletion are very small.

The effects of changing the benchmarking target are studied in a sensitivity analysis. As expected, changing the benchmarking target leads to different estimates, but, unexpectedly, the changes in the posterior standard deviations are small. Small changes in the estimates are noted for the benchmarking methods using different probabilities of deletion.

It is expected that the posterior standard deviations from deleting the last one benchmarking and random benchmarking be larger than those from the BFH model because of the jittering effect from benchmarking. However, in the empirical studies we present, deleting the last one benchmarking and random benchmarking have about the same posterior standard deviations with a small reduction when random benchmarking is used. The key strength of the random benchmarking approach is that there is no preferential treatment for any area/county.

Disclaimer and acknowledgements

The Findings and Conclusions in This Preliminary Publication Have Not Been Formally Disseminated by the U.S. Department of Agriculture and Should Not Be Construed to Represent Any Agency Determination or Policy. This research was supported in part by the intramural research program of the U.S. Department of Agriculture, National Agriculture Statistics Service.

Dr. Nandram’s work was supported by a grant from the Simons Foundation (353953, Balgobin Nandram). The authors thank the Associate Editor and the referees for their comments and suggestions. The work of Erciulescu was completed as a Research Associate at the National Institute of Statistical Sciences (NISS) working on NASS projects.

Appendix A

Exemplification of the sensitivity of deletion

Let $y_i \overset{\text{ind}}{\sim} \text{Normal}(\mu_i, \sigma_i^2)$, $i = 1, 2$, such that $y_1 + y_2 = a$, and $\lambda = \sigma_2^2/(\sigma_1^2 + \sigma_2^2)$. Then, if we start by deleting y_2, the joint density of (y_1, y_2) is

$$f(y_1, y_2 \,|\, \phi = 0) = \delta_{y2}(a - y_1)\,\text{Normal}\left\{\lambda_{\mu_1} + (1-\lambda)(a - \mu_2), (1-\lambda)\sigma_2^2\right\},$$

where $\delta_a(b) = 1$ if $a = b$ and $\delta_a(b) = 0$ if $a \neq b$. However, if we start by deleting y_1, the joint density of (y_1, y_2) is

$$g(y_1, y_2 \,|\, \phi = 0) = \delta_{y1}(a - y_2)\,\text{Normal}\{\lambda(a - \mu_1) + (1-\lambda)\mu_2, (1-\lambda)\sigma_2^2\}.$$

It will matter in the estimation procedure which variable is deleted because the two joint distributions are different. Note that the two distributions are the same if and only if

$$\lambda\mu_1 + (1-\lambda)(a - \mu_2) = \lambda(a - \mu_1) + (1-\lambda)\mu_2,$$

which gives

$$\lambda(\mu_1 - a/2) = (1-\lambda)(\mu_2 - a/2).$$

Even if we assume that $\sigma_1^2 = \sigma_2^2$, the two distributions are different. However, under this assumption, $\lambda = 1/2$, and the condition for the two distributions to be the same is that $\mu_1 = \mu_2$. That is, overall the condition for the two joint distributions to be the same is that $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$, thereby making y_1 and y_2 exchangeable. However, this is a very restricted situation.

One way out of this diffculty is to actually delete both y_1 and y_2 in the following way. Let $z = 1$ if y_1 is deleted and let $z = 0$ if y_2 is deleted. Then,

$$p(y_1, y_2, z \,|\, \phi = 0) = [pg(y_1, y_2 \,|\, \phi = 0)]^z\,[(1-p)\,f(y_1, y_2 \,|\, \phi = 0)]^{1-z},$$

where we have taken $z \sim \text{Bernoulli}(p)$ and, because z is not really identifiable, we will take $p = 1/2$ (i.e., we randomly delete one or the other). However, note that

$$z \,|\, y_1, y_2, \phi = 0 \sim \text{Bernoulli}\left\{\frac{pg(y_1, y_2 \,|\, \phi = 0)}{[pg(y_1, y_2 \,|\, \phi = 0)] + [(1-p)\,f(y_1, y_2 \,|\, \phi = 0)]}\right\}.$$

Appendix B

Fitting the Bayesian Fay-Herriot model

The Bayesian Fay-Herriot (BFH) model is given in (2.1) and the joint posterior density under the BFH model is given in (2.3), which for convenience we state here,

$$\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2 \,|\, \hat{\boldsymbol{\theta}}) \propto \frac{1}{(1+\sigma^2)^2}\left(\frac{1}{\sigma^2}\right)^{\ell/2}\prod_{i=1}^{\ell}\left\{\exp\left[-\frac{1}{2}\left\{\frac{1}{s_i^2}\left(\hat{\theta}_i - \theta_i\right)^2 + \frac{1}{\sigma^2}\left(\theta_i - \mathbf{x}_i'\boldsymbol{\beta}\right)^2\right\}\right]\right\}. \quad \text{(B.1)}$$

We show how to fit the joint posterior density of the parameters using random samples (not even a Gibbs sampler), thereby avoiding any monitoring. We will use the multiplication rule to write

$$\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2 \,|\, \hat{\boldsymbol{\theta}}) = \pi_1(\boldsymbol{\theta} \,|\, \boldsymbol{\beta}, \sigma^2, \hat{\boldsymbol{\theta}})\,\pi_2(\boldsymbol{\beta} \,|\, \sigma^2, \hat{\boldsymbol{\theta}})\,\pi_3(\sigma^2 \,|\, \hat{\boldsymbol{\theta}}),$$

where $\pi_1(\boldsymbol{\theta} \,|\, \boldsymbol{\beta}, \sigma^2, \hat{\boldsymbol{\theta}})$ and $\pi_2(\boldsymbol{\beta} \,|\, \sigma^2, \hat{\boldsymbol{\theta}})$ have standard forms and $\pi_3(\sigma^2 \,|\, \hat{\boldsymbol{\theta}})$ is nonstandard but it is density of a single parameter.

Momentarily, we will drop the term, $\frac{1}{(1+\sigma^2)^2}\left(\frac{1}{\sigma^2}\right)^{\ell/2}$, because it only affects the posterior density of σ^2. That is,

$$\pi_1\left(\boldsymbol{\theta}\,|\,\boldsymbol{\beta},\,\sigma^2,\,\hat{\boldsymbol{\theta}}\right) \propto \prod_{i=1}^{\ell}\left\{\exp\left[-\frac{1}{2}\left\{\frac{1}{s_i^2}\left(\hat{\theta}_i-\theta_i\right)^2+\frac{1}{\sigma^2}\left(\theta_i-\mathbf{x}_i'\boldsymbol{\beta}\right)^2\right\}\right]\right\}.$$

Standard calculations reduce the argument (without $-1/2$) of the exponential term to

$$\frac{1}{(1-\lambda_i)\,\sigma^2}\left\{\theta_i-\left(\lambda_i\hat{\theta}_i+(1-\lambda_i)\,\mathbf{x}_i'\boldsymbol{\beta}\right)\right\}^2+\frac{\lambda_i}{\sigma^2}\left(\hat{\theta}_i-\mathbf{x}_i'\boldsymbol{\beta}\right)^2,\ \ \lambda_i=\frac{\sigma^2}{s_i^2+\sigma^2},\ i=1,\ldots,\ell.$$

Hence, for $\pi_1\left(\boldsymbol{\theta}\,|\,\boldsymbol{\beta},\,\sigma^2,\,\hat{\boldsymbol{\theta}}\right)$,

$$\theta_i\,|\,\boldsymbol{\beta},\,\sigma^2,\,\hat{\boldsymbol{\theta}} \overset{\text{ind}}{\sim} \text{Normal}\left\{\lambda_i\hat{\theta}_i+(1-\lambda_i)\,\mathbf{x}_i'\boldsymbol{\beta},\,(1-\lambda_i)\,\sigma^2\right\},\ i=1,\ldots,\ell. \tag{B.2}$$

Momentarily, we will drop the term, $\prod_{i=1}^{\ell}[(1-\lambda_i)\,\sigma^2]^{1/2}$. Then, integrating out the θ_i, we get

$$\pi_2\left(\boldsymbol{\beta}\,|\,\sigma^2,\,\hat{\boldsymbol{\theta}}\right) \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{\ell}\frac{\lambda_i}{\sigma^2}\left(\hat{\theta}_i-\mathbf{x}_i'\boldsymbol{\beta}\right)^2\right\}.$$

Hence, the exponent (without $-1/2$) can be written as,

$$\sum_{i=1}^{\ell}\frac{\lambda_i}{\sigma^2}\left(\hat{\theta}_i-\mathbf{x}_i'\hat{\boldsymbol{\beta}}\right)^2+\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)'\hat{\Sigma}^{-1}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right),$$

where

$$\hat{\boldsymbol{\beta}}=\hat{\Sigma}\sum_{i=1}^{\ell}\frac{\hat{\theta}_i\mathbf{x}_i}{s_i^2+\sigma^2} \quad \text{and} \quad \hat{\Sigma}^{-1}=\sum_{i=1}^{\ell}\frac{\mathbf{x}_i\mathbf{x}_i'}{s_i^2+\sigma^2}.$$

It is worth noting that $\hat{\boldsymbol{\beta}}$ and $\hat{\Sigma}$ are well defined for all σ^2 provided that the design matrix, X, where $X'=(\mathbf{x}_1,\ldots,\mathbf{x}_\ell)$ is full rank. Then,

$$\pi_2\left(\boldsymbol{\beta}\,|\,\sigma^2,\,\hat{\boldsymbol{\theta}}\right) \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{\ell}\frac{\lambda_i}{\sigma^2}\left(\hat{\theta}_i-\mathbf{x}_i'\hat{\boldsymbol{\beta}}\right)^2-\frac{1}{2}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)'\hat{\Sigma}^{-1}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\right\}.$$

That is,

$$\boldsymbol{\beta}\,|\,\sigma^2,\,\hat{\boldsymbol{\theta}} \sim \text{Normal}\left(\hat{\boldsymbol{\beta}},\,\hat{\Sigma}\right). \tag{B.3}$$

Now, integrating out $\boldsymbol{\beta}$ and incorporating the terms in σ^2, which were dropped, we have

$$\pi_3\left(\sigma^2\,|\,\hat{\boldsymbol{\theta}}\right) \propto Q(\sigma^2)\frac{1}{\left(1+\sigma^2\right)^2}, \tag{B.4}$$

where

$$Q(\sigma^2)=\left|\hat{\Sigma}\right|^{1/2}\prod_{i=1}^{\ell}\frac{1}{\left(s_i^2+\sigma^2\right)^{1/2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{\ell}\frac{1}{s_i^2+\sigma^2}\left(\hat{\theta}_i-\mathbf{x}_i'\hat{\boldsymbol{\beta}}\right)^2\right\}.$$

To obtain a random sample from (B.1), we sample σ^2 from (B.4), $\boldsymbol{\beta}$ from (B.3) and the θ_i independently from (B.2). The conditional posterior density in (B.4) is nonstandard, and to draw a sample from it, we use a grid method (e.g., Nandram and Yin, 2016). First, we transform σ^2 to $\phi = \sigma^2/(1+\sigma^2)$ so that $0 < \phi < 1$. Then, we divide (0, 1) into 100 grids. Actually, we have located the range of ϕ in (0, 1) and we have divided this interval into 100 grids. This gives us a probability mass function that we sample. Jittering is used in the selected grid to get deviates, which are different with probability one; see Nandram and Yin (2016) for more details.

Appendix C

Proof of Theorem 2

It is convenient to make the following transformations,

$$\theta_i = \theta_i,\ i = 1, \ldots, \ell - 1,\ \phi = \sum_{i=1}^{\ell} \theta_i - a.$$

Here, ϕ is a dummy variable, which holds the benchmarking constraint, and it ensures a non-singular transformation. The Jacobian is unity and the inverse transformation is

$$\theta_i = \theta_i,\ i = 1, \ldots, \ell - 1,\ \theta_\ell = \phi + a - \sum_{i=1}^{\ell-1} \theta_i.$$

The transformed density is

$$\tilde{\pi}(\theta_1, \ldots, \theta_{\ell-1}, \phi).$$

Then, the density that holds the benchmarking constraint exactly is

$$\tilde{\pi}(\theta_1, \ldots, \theta_{\ell-1} \,|\, \phi = 0),\ \theta_\ell = a - \sum_{i=1}^{\ell-1} \theta_i.$$

Therefore,

$$\pi(\theta_1, \ldots, \theta_{\ell-1} \,|\, \phi = 0) \propto \exp\left\{-\frac{1}{2\delta^2}\left[\sum_{i=1}^{\ell-1}\left(\theta_i - \mathbf{u}_i'\boldsymbol{\beta}\right)^2 + \left\{\sum_{i=1}^{\ell-1}\theta_i - \left(a - \mathbf{u}_i'\boldsymbol{\beta}\right)\right\}^2\right]\right\}.$$

Dropping terms that do not involve $\boldsymbol{\theta}_{(\ell)} = (\theta_1, \ldots, \theta_{\ell-1})'$, it is easy to show that the exponent is

$$\frac{1}{2\delta^2}\left\{\boldsymbol{\theta}_{(\ell)}'(I+J)\boldsymbol{\theta}_{(\ell)} - 2\left[(\mathbf{u}_1 - \mathbf{u}_\ell)'\boldsymbol{\beta}, \ldots, (\mathbf{u}_{\ell-1} - \mathbf{u}_\ell)'\boldsymbol{\beta} + a\mathbf{j}'\right]\boldsymbol{\theta}_{(\ell)}\right\}.$$

Then, using the properties of a multivariate normal density, we have

$$\boldsymbol{\theta}_{(\ell)} \,|\, \phi = 0 \ \sim\ \text{Normal}\left((I+J)^{-1}\left(a\mathbf{j}' + (\mathbf{u}_1 - \mathbf{u}_\ell)'\boldsymbol{\beta}, \ldots, (\mathbf{u}_{\ell-1} - \mathbf{u}_\ell)'\boldsymbol{\beta}\right)', \delta^2(I+J)^{-1}\right).$$

Finally, using the Sherman-Morrison formula, $(I + J)^{-1} = I - \frac{1}{\ell} J$, we have

$$\boldsymbol{\theta}_{(\ell)} \,|\, \phi = 0 \;\sim\; \text{Normal}\left\{\left(I - \frac{1}{\ell} J\right)\left(a\mathbf{j}' + (\mathbf{u}_1 - \mathbf{u}_\ell)' \boldsymbol{\beta}, \ldots, (\mathbf{u}_{\ell-1} - \mathbf{u}_\ell)' \boldsymbol{\beta}\right)', \delta^2\left(I - \frac{1}{\ell} J\right)\right\}.$$

It is worth noting that the matrix determinant lemma gives $\det(I + J) = \ell$ and so $\det\left(\delta^2 (I - \frac{1}{\ell} J)\right) = \frac{1}{\ell}(\delta^2)^{\ell-1}$.

References

Battese, G., Harter, R. and Fuller, W. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83(401), 28-36. doi:10.2307/2288915.

Bell, W.R., Datta, G.S. and Ghosh, M. (2013). Benchmarking small area estimators. *Biometrika*, 100(1), 189-202.

Cruze, N.B., Erciulescu, A.L., Nandram, B., Barboza, W. and Young, L.J. (2019). Producing official county-level agricultural estimates in the United States: Needs and challenges. *Statistical Science*. To appear.

Datta, G.S., Ghosh, M., Steorts, R. and Maples, J. (2011). Bayesian benchmarking with applications to small area estimation. *Test*, 20(3), 574-588.

Erciulescu, A.L., Cruze, N.B. and Nandram, B. (2019). Model-based county level crop estimates incorporating auxiliary sources of information. *Journal of Royal Statistical Society, Series A*, 182, 283-303. doi:10.1111/rssa.12390.

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(366), 269-277.

Ghosh, M., and Steorts, R. (2013). Two stage Bayesian benchmarking as applied to small area estimation. *Test*, 22(4), 670-687.

Janicki, R., and Vesper, A. (2017). Benchmarking techniques for reconciling Bayesian small area models at distinct geographic levels. *Statistical Methods and Applications*, 26, 4, 557-581.

Jiang, J., and Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test*, 15(1), 1-96.

Nandram, B., and Sayit, H. (2011). A Bayesian analysis of small area probabilities under a constraint. *Survey Methodology*, 37, 2, 137-152. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2011002/article/11603-eng.pdf.

Nandram, B., and Toto, M.C.S. (2010). Bayesian predictive inference for benchmarking crop production for Iowa counties. *Journal of the Indian Society of Agricultural Statistics*, 64 (2), 191-207.

Nandram, B., and Yin, J. (2016). A nonparametric Bayesian prediction interval for a finite population mean. *Journal of Statistical Computation and Simulation*, 86 (16), 3141-3157.

Nandram, B., Berg, E. and Barboza, W. (2014). A hierarchical Bayesian model for forecasting state-level corn yield. *Journal of Environmental and Ecological Statistics*, 21, 507-530.

Nandram, B., Toto, M.C.S. and Choi, J.W. (2011). A Bayesian benchmarking of the Scott-Smith model for small areas. *Journal of Statistical Computation and Simulation*, 81 (11), 1593-1608.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, 28(1), 40-68.

Pfeffermann, D., Sikov, A. and Tiller, R. (2014). Single-and two-stage cross-sectional and time series benchmarking procedures for small area estimation. *Test*, 23(4), 631-666.

Pfeffermann, D., and Tiller, R. (2006). Small area estimation with state-space models subject to benchmark constraints. *Journal of the American Statistical Association*, 101 (476), 1387-1397.

Rao, J.N.K., and Molina, I. (2015). *Small Area Estimation*, Wiley Series in Survey Methodology.

Toto, M.C.S., and Nandram, B. (2010). A Bayesian predictive inference for small area means incorporating covariates and sampling weights. *Journal of Statistical Planning and Inference*, 140, 2963-2979.

Wang, J., Fuller, W.A. and Qu, Y. (2008). Small area estimation under a restriction. *Survey Methodology*, 34, 1, 29-36. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2008001/article/10619-eng.pdf.

You, Y., and Rao, J.N.K. (2002). Small area estimation using unmatched sampling and linking models. *The Canadian Journal of Statistics*, 30, 3-15.

You, Y., Rao, J.N.K. and Dick, J.P. (2004). Benchmarking hierarchical Bayes small area estimators in the Canadian census under coverage estimation. *Statistics in Transition*, 6, 631-640.