**United States Department of Agriculture**

**National Agricultural Statistics Service**

**Research and Development Division Washington DC 20250**

**RDD Research Report Number RDD-22-02**

**July 2022**

# Bayesian Small Area Models under Inequality Constraints with Benchmarking and Double Shrinkage

Balgobin Nandram
Nathan B. Cruze
Andreea L. Erciulescu
Lu Chen

# TABLE OF CONTENTS

# Bayesian Small Area Models under Inequality Constraints with Benchmarking and Double Shrinkage

Balbogin Nandram[*], Nathan B. Cruze[†], Andreea L. Erciulescu[‡], Lu Chen[§]

**Abstract**

We present a novel methodology to benchmark county-level estimates of crop area totals to a preset state total subject to inequality constraints and random variances in the Fay-Herriot model. For planted area of the United States Department of Agriculture, it is necessary to incorporate the constraint that the estimated totals, derived from survey and other auxiliary data, are no smaller than administrative planted area totals prerecorded by other agencies. These administrative totals are treated as fixed and known, and this additional coherence requirement adds to the complexity of benchmarking the county-level estimates. A fully Bayesian analysis of the Fay-Herriot model offers an appealing way to incorporate the inequality and benchmarking constraints, and to quantify the resulting uncertainties, but sampling from the posterior densities involves difficult integration, and reasonable approximations must be made. First, we describe a single-shrinkage model, shrinking the means while the variances are assumed known. Second, we extend this model to accommodate double shrinkage, borrowing strength across means and variances. This extended model has two sources of extra variation, but because we are shrinking both means and variances, it is expected that this second model should perform better in terms of precision and goodness of fit. Both models are applied to simulated data sets with properties resembling the Illinois corn crop.

**KEY WORDS**: Devroye method; Fay-Herriot model; Grid method; Hierarchical Bayesian model; Metropolis sampler.

[*]Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609 and USDA National Agricultural Statistical Service, 1400 Independence Avenue, Washington, DC 20250-2054

[†]NASA Langley Research Center, Mail Stop 290, Hampton, VA 23681

[‡]Westat, 1600 Research Blvd., Rockville, MD 20850

[§]National Institute of Statistical Sciences,1750 K Street, NW, Suite 1100 Washington, DC 20006 and USDA National Agricultural Statistical Service, 1400 Independence Avenue, Washington, DC 20250-2054

# 1 INTRODUCTION

For many problems in official statistics, it is necessary to incorporate constraints in model-based inference. For example, in small area estimation, there may be constraints on the model estimates, which are to be benchmarked to a target. These may be known lower (or upper) bounds for county estimates, which should "add up" to the state estimate, obtained earlier. One practical example is the estimation of planted acres for counties within states, with a state estimate obtained earlier, when there are survey data and administrative data that can provide lower bounds to the county estimates, which are required to add up to the state estimate. While we focus on an application in agriculture, we develop a methodology to solve the problem in which small area estimates are needed to satisfy certain lower bounds and these estimates are further benchmarked to an estimate at a higher level via the top down approach.

In the United States, official county-level estimates of crop yield, total production, and total acreage published by United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) are important. These official estimates may determine the amount of payments to be made to farmers and ranchers enrolled in several programs administered by other USDA agencies including the Farm Service Agency (FSA) and the Risk Management Agency (RMA). Accordingly, NASS strives to improve the accuracy, reliability, and coverage of its official crop county estimates. As described in a report titled *Improving Crop County Estimates by Integrating Multiple Data Sources* (National Academies of Sciences, Engineering, and Medicine 2017), one way to do so is to use defensible models that include multiple sources of variability and other auxiliary data. The report highlighted many of the challenges faced by NASS and emphasized the role model-based inference can play in the publication of official county estimates. The findings of the report were further discussed in Cruze et al. (2019), and the authors identified coherence of crop area estimates with known, same-year administrative acreage totals as a significant need for the NASS crops county estimates program.

Constraints on estimates may enter in the form of order or shape restrictions (e.g., Nandram, Sedransk and Smith 1997; Silvapulle and Sen 2005; Chen and Nandram 2022) or in the form of inequality constraints (Sen and Silvapulle 2002). The latter type of restriction is of particular interest as it relates to the coherence of tabulated crop estimates in the presence of available

administrative data curated by USDA. Benchmarking estimates for smaller geographic domains to those of larger geographic areas is one common form of equality constraint encountered in official statistics. For example, several past NASS studies have achieved this by ratio adjustment (raking) made after model output analysis (e.g., Erciulescu, Cruze, and Nandram 2018, 2019, 2020); see also Steorts, Schmid and Tzavidis (2020) and the references therein for an informative review on benchmarking. While the emphasis of the present work is methodological, we note that a recent NASS-authored case study and companion paper (Chen, Cruze, and Nandram 2022) on the constrained planted area problem, single shrinkage model, will appear in a forthcoming issue in the *Journal of Official Statistics*, adding to the growing literature on statistical inference under inequality constraints. Also, we note that our main contributions are on the inequality constraints.

Non-probability data are not devoid of errors. First, it is understood that while participation in agricultural support programs is popular in the United States, the voluntary enrollment in FSA and RMA programs contributes to potential under-coverage (a downward bias) in these administrative acreage totals. Moreover, rates of participation in these support programs may differ each year, by commodity crop, by state, or even more locally within state. Other nonsampling errors, however, are believed to be minimized through FSA and RMA quality controls. For example, farmers certify their enrolled acreages with FSA agents on geolocated field boundaries, and farmers are subject to penalties for falsifying their reports. With these properties in mind, the available administrative totals are viewed by NASS and USDA as *informative lower bounds* and publication of coherent tabular data on planted area requires: 1) that county acreage totals sum to the state acreage totals that are published prior to the release of county estimates, and 2) that official county-level planted area estimates honor the lower bound constraint in each county.

Additionally, we consider possible gains from double shrinkage by borrowing strength from means and variances simultaneously. Both frequentist and Bayesian model-based estimation techniques for the sampling variances have been considered in the literature for the area-level models. For example, see Wang and Fuller (2003); You and Chapman (2006); Gonzalez-Manteiga et al. (2010), Erciulescu and Berg (2014); Maiti, Ren and Sinha (2014); and Dass et al. (2012). Recently, Erciulescu, Cruze and Nandram (2019) incorporated double shrinkage in estimates of unconstrained harvested area totals.

Let $\hat{\theta}_i, i = 1, \ldots, \ell$, denote the observed direct estimates of total acreage for $\ell$ counties, and

$\hat{\sigma}_i^2, i = 1, \ldots, \ell$, denote the corresponding observed variances for the $\ell$ counties; to avoid hat notations when fractions are written, sometimes we prefer to use $S_i^2$ instead of $\hat{\sigma}_i^2$. The Fay-Herriot (area-level) model (Fay and Herriot 1979; see also, Rao and Molina 2015) is a standard model in small area estimation for the $\hat{\theta}_i$ where:

$$\hat{\theta}_i \mid \theta_i \overset{ind}{\sim} \text{Normal}(\theta_i, \hat{\sigma}_i^2), i = 1, \ldots, \ell, \tag{1}$$

and at the second stage,

$$\theta_i \mid \underset{\sim}{\beta}, \delta^2 \overset{ind}{\sim} \text{Normal}(\underset{\sim}{x}_i'\underset{\sim}{\beta}, \delta^2), i = 1, \ldots, \ell, \tag{2}$$

where $\underset{\sim}{x}_i$ is a $p$-vector of covariates with an intercept and $\underset{\sim}{\beta}$ is a $p$-vector of regression coefficients. In a full Bayesian analysis of this model, prior distributions of model parameters are assumed; a priori we take $\pi(\underset{\sim}{\beta}, \delta^2) = \pi(\underset{\sim}{\beta})\pi(\delta^2)$, where $\pi(\delta^2)$ is proper but $\pi(\underset{\sim}{\beta}) = 1$ is improper.

Procedurally, NASS state estimates of planted area (denote these state targets by the scalar $a$) are determined and published prior to the publication of county-level estimates. Nandram, Erciulescu and Cruze (2019) developed a full Bayesian Fay-Herriot model incorporating the benchmarking constraint $\sum_{i=1}^{\ell} \theta_i = a$ directly into the model. This was achieved by deleting the last area to accommodate the benchmarking constraint. They empirically showed that, in practice, it does not really matter much which area is deleted in order to incorporate the benchmarking constraint.

We now want to refine this model to accommodate benchmarking and inequality constraints on the $\theta_i$. In addition to the benchmarking constraint, we need to add the county-specific inequality constraints

$$\theta_i \geq c_i, i = 1, \ldots, \ell, \tag{3}$$

where the $c_i$ are fixed, known quantities that represent administrative values provided by FSA or RMA. (In practice, when both data sources are present, the larger of the two is used to establish the lower bound, $c_i$.) In NASS planted acres data, some of the direct estimates of planted area totals may be more than one or two standard errors below their corresponding $c_i$, thereby creating some difficulties for the model estimates to be larger than the $c_i$. It is worth noting that

$a = \sum_{i=1}^{\ell} \theta_i \geq \sum_{i=1}^{\ell} c_i \equiv c$. That is, the estimation processes that generate state targets also respect the available administrative totals at *state level*, however, the benchmarking constraint can create additional difficulties when the target is only slightly larger than $c$, i.e., as $\frac{c}{a} \to 1$ from below. We need to add the inequality constraints to the Fay-Herriot model specified in (1), (2) *and the priors* to get the joint posterior density of $\theta_i, i = 1, \ldots, \ell$. In order to incorporate the inequality constraints into the Bayesian Fay-Herriot model, we propose the following simplification. In departure from Nandram, Erciulescu and Cruze (2019), we incorporate the inequality constraints directly while only partially incorporating the benchmarking constraint into the Bayesian Fay-Herriot model. That is, we will incorporate the constraints, $c_i \leq \theta_i, i = 1, \ldots, \ell$, together with the restriction that $\sum_{i=1}^{\ell} \theta_i < a$ into the model. When the latter inequality is enforced, a raking of model estimates to the state total $a$ in an output analysis will still satisfy all individual county inequality constraints. Incorporating double shrinkage into the inequality-constrained model entails additional computational considerations. Therefore, our key contributions are to provide small area estimates, which are subjected to inequality constraints, benchmarked to a target, and we describe a single shrinkage model (sample variances fixed) and two double shrinkage models (sample variances random).

In this paper, we discuss a novel methodology to solve these dual problems by modifying the Bayesian Fay-Herriot model described in Nandram, Erciulescu, and Cruze (2019) to accommodate both benchmarking and inequality constraints into the Bayesian area-level models of Equations (1) and (2). Additionally, we extend the model to accommodate double shrinkage of means and variances. In Section 2, we introduce the methodology for single-shrinkage model in the presence of inequality constrained totals. In Section 3, we describe the methodology for the double-shrinkage model, gamma regression, and the log-linear model is discussed in Appendix B; again double-shrinkage models incorporate inequality constrained totals. Special emphasis is given to the computation that facilitates these approaches. In Section 4, as confidentiality of USDA survey and administrative data is a concern, simulated data sets with properties resembling those of the Illinois corn crop are generated and used to fit and assess these models. We offer concluding remarks in Section 5, noting that constrained acreage methodologies were successfully incorporated in NASS official statistics beginning with the 2020 crop year.

## 2 METHODOLOGY UNDER THE SINGLE SHRINKAGE MODEL

In this section, we develop the methodologies and computational strategies to incorporate inequality constraints and benchmarking procedures into the Bayesian area-level models of Equations (1) and (2). This provides the single shrinkage model in which the sampling variances are assumed fixed and known.

Our strategy is to use the composition rule (i.e., multiplication rule of probability) to draw samples from the posterior density $\pi(\underset{\sim}{\beta}, \delta^2 \mid \underset{\sim}{\hat{\theta}}, \underset{\sim}{\hat{\sigma}}^2)$ and then to draw samples from $\pi(\underset{\sim}{\theta} \mid \underset{\sim}{\beta}, \delta^2, \underset{\sim}{\hat{\theta}}, \underset{\sim}{\hat{\sigma}}^2)$. Both of these problems are difficult. In this section, we have used the shrinkage prior for $\delta^2$ (i.e., $\pi(\delta^2) = 1/(1+\delta^2)^2, \delta^2 > 0$) to avoid impropriety of the posterior density. Letting $\phi = 1/(1+\delta^2)^2$, then $\phi \sim \text{Beta}(1,1)$ (i.e., uniform) and this is similar to the half Cauchy prior $\pi(\delta^2) = \frac{1}{\pi\sqrt{\delta^2}(1+\delta^2)}$, which translates to $\phi \sim \text{Beta}(.5,.5)$. In addition, both densities are in the Snedecor $f$ form, where the first density is a $f(2,2)$ and the Cauchy version is a $f(1,1)$; the $f(2,2)$ is mathematically a bit more convenient when we transform to $(0,1)$.

Let $V = \{\underset{\sim}{\theta} : \theta_i \geq c_i, i = 1, \ldots, \ell, \sum_{i=1}^{\ell} \theta_i < a\}$. Here, this conditional posterior density, $\pi(\underset{\sim}{\theta} \mid \underset{\sim}{\beta}, \delta^2, \underset{\sim}{\hat{\theta}}, \underset{\sim}{\hat{\sigma}}^2)$, is subject to the inequality constraint and the constraint $\sum_{i=1}^{\ell} \theta_i < a$, where $a$ is the benchmarking target. Note that the inequality is strict because with the equality, one of the $\theta_i$ becomes redundant. This redundancy has to be taken into consideration when the model is fit (a much more difficult problem), but with the inequality constraint we do not need to do so (a much easier problem). That is, we need to draw $\theta_1, \ldots, \theta_\ell$ subject to the constraints $\theta_i \geq c_i, i = 1, \ldots, \ell$ and $\sum_{i=1}^{\ell} \theta_i < a$. Note again that the benchmarking constraint is only partially included in the Fay-Herriot model. We will use a Gibbs sampler to carry out this sampling procedure, and the benchmarking constraint will be fully incorporated in an output analysis from the Gibbs sampler using a raking procedure.

The joint prior density is

$$\pi(\underset{\sim}{\theta}, \underset{\sim}{\beta}, \delta^2) = \pi(\underset{\sim}{\beta}, \delta^2) \frac{\prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})/\delta\}/\delta}{\int_{\underset{\sim}{\theta} \in V} \prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})/\delta\}/\delta \; d\underset{\sim}{\theta}}, \; \underset{\sim}{\theta} \in V, \tag{4}$$

where $\phi(\cdot)$ is the standard normal density. Indeed, this is a very awkward joint prior density with the normalization constant a function of $(\underset{\sim}{\beta}, \delta^2)$. Then, using Bayes' theorem, the joint posterior

density is

$$\pi(\underset{\sim}{\theta}, \underset{\sim}{\beta}, \delta^2 \mid \hat{\underset{\sim}{\theta}}, \hat{\underset{\sim}{\sigma}}^2) \propto \pi(\underset{\sim}{\beta}, \delta^2) \frac{\prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})/\delta\}}{\int_{\underset{\sim}{\theta} \in V} \prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})/\delta\}\ d\underset{\sim}{\theta}} [\prod_{i=1}^{\ell} \phi\{(\theta_i - \hat{\theta}_i)/\hat{\sigma}_i\}],\ \underset{\sim}{\theta} \in V. \tag{5}$$

It is difficult to use Markov chain Monte Carlo methods to efficiently draw samples from $\pi(\underset{\sim}{\theta}, \underset{\sim}{\beta}, \delta^2 \mid \hat{\underset{\sim}{\theta}}, \hat{\underset{\sim}{\sigma}}^2)$ in (5).

We now show how to draw samples from $\pi(\underset{\sim}{\theta}, \underset{\sim}{\beta}, \delta^2 \mid \hat{\underset{\sim}{\theta}}, \hat{\underset{\sim}{\sigma}}^2)$ using numerical integration, the Gibbs sampler and the Metropolis sampler. [Note that in the discussion below, apart from $\sum_{i=1}^{\ell} \theta_i < a$, it does not matter whether we use "less than or equal" symbols because the $\theta_i$ are continuous random variables.]

We first show how to draw the $\theta_i$ using the Gibbs sampler. For the constraints, we have $c_i \leq \theta_i, i = 1, \ldots, \ell$, and $\sum_{i=1}^{\ell} \theta_i < a$. This means that $\sum_{i=1}^{\ell} c_i < \sum_{i=1}^{\ell} \theta_i < a$, and so $\max(c_i, \sum_{j=1}^{\ell} c_j - \sum_{j=1, j\neq i}^{\ell} \theta_j) < \theta_i < a - \sum_{j=1, j\neq i}^{\ell} \theta_j, i = 1, \ldots, \ell$. Therefore, the support of the conditional posterior density of $\theta_i$ given $\underset{\sim}{\theta}_{(i)} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_\ell)'$, is

$$\max(c_i, \sum_{j=1}^{\ell} c_j - \sum_{j=1, j\neq i}^{\ell} \theta_j) < \theta_i < a - \sum_{j=1, j\neq i}^{\ell} \theta_j, i = 1, \ldots, \ell.$$

It is easy to show that the conditional posterior density is

$$\theta_i \mid \underset{\sim}{\theta}_{(i)}, \underset{\sim}{\beta}, \delta^2, \hat{\underset{\sim}{\theta}}, \hat{\underset{\sim}{\sigma}}^2 \sim \text{Normal}\{\lambda_i \hat{\theta}_i + (1 - \lambda_i) \underset{\sim}{x}_i'\underset{\sim}{\beta}, (1 - \lambda_i)\delta^2\}, \lambda_i = \delta^2/(\delta^2 + \hat{\sigma}_i^2),$$

$$u_i = \max(c_i, \sum_{j=1}^{\ell} c_j - \sum_{j=1, j\neq i}^{\ell} \theta_j) < \theta_i < a - \sum_{j=1, j\neq i}^{\ell} \theta_j = v_i, i = 1, \ldots, \ell. \tag{6}$$

Now, we want to draw $\theta_i$ subject to the constraint, $u_i \leq \theta_i \leq v_i$. To sample $X \sim \text{Normal}(\mu, \sigma^2), a \leq X \leq b$, we have the following result (see Devroye 1986),

$$X = \mu + \sigma \Phi^{-1} \left\{ (1 - U)\Phi(\frac{a - \mu}{\sigma}) + U\Phi(\frac{b - \mu}{\sigma}) \right\},$$

where $U \sim \text{Uniform}(0,1)$ and $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ are respectively the cdf and the inverse cdf of the standard normal density. We use the Gibbs sampler to draw a sample $\underset{\sim}{\theta}$ in (6). This is obtained by drawing $u_i \leq \theta_i \leq v_i, i = 1, \ldots, n$, each in turn.

The final step is to rake up $\theta_1, \ldots, \theta_\ell$ to the target $a$ for each iterate. So that the final iterates are

$$\tilde{\theta}_i = \frac{a}{\sum_{j=1}^{\ell} \theta_j} \theta_i, i = 1, \ldots, \ell,$$

and posterior inference can be made about $\theta_1, \ldots, \theta_\ell$ using these raked vectors of iterates. It is now clear why $\sum_{i=1}^{\ell} < a$. Note again that this is a straight forward output analysis from the Gibbs sampler.

We next show how to draw samples from $\pi(\underset{\sim}{\beta}, \delta^2 \mid \underset{\sim}{\hat{\theta}}, \underset{\sim}{\hat{\sigma}}^2)$ using numerical integration and the Metropolis sampler. The joint posterior density of $(\underset{\sim}{\beta}, \delta^2)$ is

$$\pi(\underset{\sim}{\beta}, \delta^2 \mid \underset{\sim}{\hat{\theta}}, \underset{\sim}{\hat{\sigma}}^2) \propto \pi(\underset{\sim}{\beta}, \delta^2) \frac{\int_{\underset{\sim}{\theta} \in V} \prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i' \underset{\sim}{\beta})/\delta\} \phi\{(\theta_i - \hat{\theta}_i)/\hat{\sigma}_i\} \; d\underset{\sim}{\theta}}{\int_{\underset{\sim}{\theta} \in V} \prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i' \underset{\sim}{\beta})/\delta\} \; d\underset{\sim}{\theta}},$$

which, by completing the squares, can be simplified to

$$\pi(\underset{\sim}{\beta}, \delta^2 \mid \underset{\sim}{\hat{\theta}}, \underset{\sim}{\hat{\sigma}}^2) \propto \pi(\underset{\sim}{\beta}, \delta^2) [\prod_{i=1}^{\ell} \phi\{(\hat{\theta}_i - \underset{\sim}{x}_i' \underset{\sim}{\beta})/\sqrt{\delta^2/\lambda_i}\}] R(\underset{\sim}{\beta}, \delta^2), \tag{7}$$

with

$$R(\underset{\sim}{\beta}, \delta^2) = \frac{\int_{\underset{\sim}{\theta} \in V} \prod_{i=1}^{\ell} \phi\{(\theta_i - \mu_i)/\tau_i\} \; d\underset{\sim}{\theta}}{\int_{\underset{\sim}{\theta} \in V} \prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i' \underset{\sim}{\beta})/\delta\} \; d\underset{\sim}{\theta}},$$

where $\mu_i = \lambda_i \hat{\theta}_i + (1 - \lambda_i) \underset{\sim}{x}_i' \underset{\sim}{\beta}$ and $\tau_i^2 = (1 - \lambda_i)\delta^2$, $i = 1, \ldots, \ell$. We will use the Metropolis sampler to fit (7). There are two key issues, which are to construct an efficient proposal density and to compute the ratio, $R(\underset{\sim}{\beta}, \delta^2)$, of the two integrals in (7).

First, we consider how to construct a proposal density. We have samples of $(\underset{\sim}{\beta}, \delta^2)$ from the Fay-Herriot model. We can now transform $\delta^2$ to $\beta_{p+1} = \log(\delta^2)$ and add it as the last component to get a new vector $\underset{\sim}{\beta}$ with $p+1$ components. Now fit a multivariate normal density to the samples, $\underset{\sim}{\beta} \sim \text{Normal}(\underset{\sim}{\hat{\beta}}, \sigma^2 \hat{\Sigma})$, where $\underset{\sim}{\hat{\beta}}$ and $\hat{\Sigma}$ are the posterior mean and covariance matrix of the samples

from the Fay-Herriot model, and $\eta/\sigma^2 \sim \text{Gamma}(\eta/2, 1/2)$ to complete the $(p+1)$-variate Student's $t$ density on $\eta$ degrees of freedom, where $\eta$ is a tuning constant.

Second, we describe how to estimate the ratio of the integrals in (7). Let $\tilde{V} = \{\underset{\sim}{\theta} : c_i < \theta_i < \infty, i = 1, \ldots, \ell\}$; we have actually selected an upper bound for each $\theta_i$. Note that $V \subset \tilde{V}$, and perhaps $\tilde{V}$ is not much bigger than $V$. Let $I(\underset{\sim}{\theta} \in V) = 1$ if $\underset{\sim}{\theta} \in V$ and $I(\underset{\sim}{\theta} \in V) = 0$ otherwise. Then,

$$R(\underset{\sim}{\beta}, \delta^2) = \frac{\displaystyle\int_{\underset{\sim}{\theta} \in \tilde{V}} I(\underset{\sim}{\theta} \in V) \prod_{i=1}^{\ell} \phi\{(\theta_i - \mu_i)/\tau_i\} \, d\underset{\sim}{\theta}}{\displaystyle\int_{\underset{\sim}{\theta} \in \tilde{V}} I(\underset{\sim}{\theta} \in V) \prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})/\delta\} \, d\underset{\sim}{\theta}}.$$

Now, $R(\underset{\sim}{\beta}, \delta^2)$ can be calculated using Monte Carlo methods. As an importance function, we use the conditional posterior densities of the $\theta_i, i = 1, \ldots, \ell$, constrained on $\tilde{V}$. That is,

$$\theta_i \mid \underset{\sim}{\beta}, \delta^2 \overset{ind}{\sim} \text{Normal}(\mu_i, \tau_i^2), \; c_i < \theta_i < \infty, i = 1, \ldots, \ell. \tag{8}$$

It is now easy to draw samples $\underset{\sim}{\theta}^{(h)}, h = 1, \ldots, M$, in (8), where $M \approx 1000$ or so; see Devroye (1986). Then, a Monte Carlo estimator of $R(\underset{\sim}{\beta}, \delta^2)$ is

$$\widehat{R(\underset{\sim}{\beta}, \delta^2)} = \frac{\displaystyle\sum_{h=1}^{M} I(\underset{\sim}{\theta}^{(h)} \in V)}{\displaystyle\sum_{h=1}^{M} I(\underset{\sim}{\theta}^{(h)} \in V) \left[ \prod_{i=1}^{\ell} \frac{\phi\{(\theta_i^{(h)} - \underset{\sim}{x}_i'\underset{\sim}{\beta})/\delta\}}{\phi\{(\theta_i^{(h)} - \mu_i)/\tau_i\}} \right]}.$$

Note that for each $h$, once $\theta_i^{(h)}, i = 1, \ldots, \ell$, are drawn from the proposal density, we simply need to check that $\sum_{i=1}^{\ell} \theta_i^{(h)} < a$. However, it is possible that this Monte Carlo estimator does not exist, and this clearly occurs when $\underset{\sim}{\theta}^{(h)} \notin V, h = 1, \ldots, M$ (all $M$), and in this case we use the modified estimator,

$$\widehat{R_m(\underset{\sim}{\beta}, \delta^2)} = \left[ \frac{1}{M} \sum_{h=1}^{M} \prod_{i=1}^{\ell} \frac{\phi\{(\theta_i^{(h)} - \underset{\sim}{x}_i'\underset{\sim}{\beta})/\delta\}}{\phi\{(\theta_i^{(h)} - \mu_i)/\tau_i\}} \right]^{-1}.$$

That is, we simply replace $V$ by $\tilde{V}$ to form an approximation in the case that the Monte Carlo estimator might not exist. In either case, we have drawn the $\theta_i$ as in (8), where $\theta_i \mid \underset{\sim}{\beta}, \delta^2 \overset{ind}{\sim} \text{Normal}(\mu_i, \tau_i^2), \; c_i < \theta_i < \infty, i = 1, \ldots, \ell$. It is possible for some of the $\underset{\sim}{\theta}^{(h)}$ to be in $V$, and in this

case if the number of $\underset{\sim}{\theta}^{(h)} \in V$ is at least $M/2$, we use the former estimator.

Our procedure gives us $1,000$ samples from the posterior density of $(\underset{\sim}{\beta}, \delta^2)$ using the Metropolis sampler. Then the more important samples of $\theta_1, \ldots, \theta_\ell$ are obtained using the Gibbs sampler. For each of the $1,000$ iterates of $(\underset{\sim}{\beta}, \delta^2)$ from the Metropolis sampler, we run the Gibbs sampler to say, 100 iterations or so, and pick the last set of $\theta_1, \ldots, \theta_\ell$. This is not too expensive and it is reasonably efficient. In this method, it is not really necessary to monitor the Gibbs sampler for convergence because we need only one value but a "burn-in" is required.

# 3 METHODOLOGY UNDER THE DOUBLE SHRINKAGE MODEL

Two double shrinkage models are introduced, where we model both the sample variances and the means. The inequality constraints are also included. Here borrowing of strength occurs via both the means and the variances. For the specification of variances, the first uses a gamma regression model and the second uses a log-linear model. In Section 3, we model the sample variances using gamma regression; Section 3.1 describes the method and Section 3.2 describes the computation; further computations are shown in Appendix A. In Appendix B, we describe the second double shrinkage model for the sample variances using the log-linear model. Even a full Bayesian treatment of the log-linear model offers remarkable computational advantages relative to the gamma regression model.

## 3.1 Gamma Regression Model

For $\ell$ areas, we have the survey estimates $\hat{\theta}_i$, their standard errors $S_i$, and the sample sizes $n_i \geq 2$ (sample sizes must be at least 2). We start with a convenient model that builds upon our work on the Fay-Herriot model. We assume that

$$\hat{\theta}_i \mid \theta_i, \sigma_i^2 \overset{ind}{\sim} \text{Normal}(\theta_i, \sigma_i^2), i = 1, \ldots, \ell,$$

$$\frac{(n_i - 1)S_i^2}{\sigma_i^2} \mid \sigma_i^2 \overset{ind}{\sim} \text{Gamma}(\frac{n_i - 1}{2}, \frac{1}{2}), i = 1, \ldots, \ell,$$

where $X \sim \text{Gamma}(a, b)$ means that $f(x) = b^a x^{a-1} e^{-bx}/\Gamma(a), x \geq 0$. Note that we are assuming $\hat{\theta}_i$ and $S_i^2$ are independent. Under the first assumption, the $\theta_i$ and $\sigma_i^2$ are not estimable, but the first

and second assumptions together make $\theta_i$ and $\sigma_i^2$ estimable. A priori, we assume that

$$\theta_i \mid \underset{\sim}{\beta}, \delta^2 \overset{ind}{\sim} \text{Normal}(\underset{\sim}{x}_i'\underset{\sim}{\beta}, \delta^2), i = 1, \ldots, \ell,$$

$$\sigma_i^{-2} \mid \alpha, \underset{\sim}{\gamma} \overset{ind}{\sim} \text{Gamma}(\frac{\alpha}{2}, \frac{\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\gamma}}}{2}), i = 1, \ldots, \ell.$$

These assumptions on $\theta_i$ and $\sigma_i^2$ provide double shrinkage (shrinking both means and variances). Here, we have assumed that the two sets of covariates are the same, but they can, of course, be different. It is worth noting that the prior for $\sigma_i^2$ is conjugate providing some simplicity in the computations; see Nandram and Erhardt (2004) for similar specifications for the corresponding binomial and Poisson models. Our prior for the hyperparameters is

$$\pi(\underset{\sim}{\beta}, \delta^2, \underset{\sim}{\gamma}, \alpha) \propto \frac{1}{(1+\delta^2)^2} \frac{1}{(1+\alpha)^2}, \delta^2, \alpha \geq 0.$$

That is, flat priors are assumed for $\underset{\sim}{\beta}$ and $\underset{\sim}{\gamma}$, shrinkage priors (proper) are assumed on $\delta^2$ and $\alpha$, and all parameters are independent. Note that $\delta^2$ and $\alpha$ are nonnegative, and so we prefer to use a shrinkage prior. At this point, there are virtually no mathematical, computational or scientific benefits using other noninformative priors for $\alpha$.

In our model, we include the inequality constraint, $\theta_i > c_i, i = 1, \ldots, \ell, \sum_{i=1}^{\ell} \theta_i < a$, where $a$ is the target. Note again that we only partially include the benchmarking constraint. It is convenient that this is the same region as for single shrinkage model, $V = \{\underset{\sim}{\theta} : \theta_i \geq c_i, i = 1, \ldots, \ell, \sum_{i=1}^{\ell} \theta_i < a\}$. Therefore, the prior densities for the $\theta_i$ remain the same,

$$\pi(\underset{\sim}{\theta}, \underset{\sim}{\beta}, \delta^2) = \pi(\underset{\sim}{\beta}, \delta^2) \frac{\prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})/\delta\}}{\int_{\underset{\sim}{\theta} \in V} \prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})/\delta\} \, d\underset{\sim}{\theta}}, \ \underset{\sim}{\theta} \in V,$$

where $\phi(\cdot)$ is the standard normal density. It is convenient to define $\Omega = (\underset{\sim}{\beta}, \delta^2, \underset{\sim}{\gamma}, \alpha)$. Then, the joint prior density is

$$\pi(\underset{\sim}{\theta}, \underset{\sim}{\sigma}^2, \Omega) = \pi(\Omega) \frac{\prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})/\delta\}}{\int_{\underset{\sim}{\theta} \in V} \prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})/\delta\} \, d\underset{\sim}{\theta}}$$

$$\times \prod_{i=1}^{\ell} \left\{ (\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\gamma}}/2)^{\alpha/2} (1/\sigma_i^2)^{\alpha/2+1} e^{-(\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\gamma}}/2\sigma_i^2)} / \Gamma(\alpha/2) \right\}, \ \underset{\sim}{\theta} \in V. \tag{9}$$

By independence, the joint density of $(\hat{\underset{\sim}{\theta}}, \underset{\sim}{S}^2)$, is

$$f(\hat{\underset{\sim}{\theta}}, \underset{\sim}{S}^2 \mid \underset{\sim}{\theta}, \underset{\sim}{\sigma}^2, \Omega) = \prod_{i=1}^{\ell} \left\{ \frac{1}{\sigma_i} \phi\{ \frac{\hat{\theta}_i - \theta_i}{\sigma_i} \} \right.$$

$$\times \prod_{i=1}^{\ell} \left\{ \{ [(n_i - 1)/2\sigma_i^2]^{(n_i-1)/2} (s_i^2)^{(n_i-1)/2-1} e^{-(n_i-1)s_i^2/2\sigma_i^2} \} / \Gamma\{(n_i - 1)/2\} \right\}. \tag{10}$$

Finally, using Bayes' theorem, the joint posterior density is proportional to the product of (9) and (10) and it can be shown to be

$$\pi(\underset{\sim}{\theta}, \underset{\sim}{\sigma}^2, \Omega \mid \hat{\underset{\sim}{\theta}}, \underset{\sim}{S}^2) \propto \pi(\underset{\sim}{\beta}, \delta^2, \underset{\sim}{\gamma}, \alpha) \frac{1}{\int_{\underset{\sim}{\theta} \in V} \prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})/\delta\} \ d\underset{\sim}{\theta}}$$

$$\times \prod_{i=1}^{\ell} \left\{ (\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\gamma}}/2)^{\alpha/2} (1/\sigma_i^2)^{\alpha/2+1} e^{-(\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\gamma}}/2\sigma_i^2)} / \Gamma(\alpha/2) \right\}$$

$$\times \prod_{i=1}^{\ell} \left\{ \frac{1}{\sqrt{(1-\lambda_i)\delta^2}} \phi \left( \frac{\theta_i - (\lambda_i \hat{\theta}_i + (1-\lambda_i)\underset{\sim}{x}_i'\underset{\sim}{\beta})}{\sqrt{(1-\lambda_i)\delta^2}} \right) \frac{1}{\sqrt{\delta^2/\lambda_i}} \phi( \frac{\hat{\theta}_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}}{\sqrt{\delta^2/\lambda_i}} ) \right\}$$

$$\times \prod_{i=1}^{\ell} \left\{ [(n_i - 1)/2\sigma_i^2]^{(n_i-1)/2} e^{-(n_i-1)s_i^2/2\sigma_i^2} \right\}, \ \underset{\sim}{\theta} \in V, \tag{11}$$

where $\lambda_i = \delta^2/(\delta^2 + \sigma_i^2), i = 1, \ldots, \ell$.

It now follows from (11) that the conditional posterior densities of the $\theta_i$ are

$$\theta_i \mid \underset{\sim}{\sigma}^2, \Omega, \hat{\underset{\sim}{\theta}}, \underset{\sim}{S}^2 \overset{ind}{\sim} \text{Normal}\{\lambda_i \hat{\theta}_i + (1-\lambda_i)\underset{\sim}{x}_i'\underset{\sim}{\beta}, (1-\lambda_i)\delta^2\}, i = 1, \ldots, \ell, \ \underset{\sim}{\theta} \in V. \tag{12}$$

Now, one can integrate out the $\theta_i$ from (11) to get the joint conditional posterior density of $\underset{\sim}{\sigma}^2$,

$$\pi(\underset{\sim}{\sigma}^2 \mid \Omega, \underset{\sim}{\hat{\theta}}, \underset{\sim}{S}^2) \propto \int_{\underset{\sim}{\theta} \in V} \prod_{i=1}^{\ell} \phi\{\frac{\theta_i - (\lambda_i \hat{\theta}_i + (1-\lambda_i)\underset{\sim}{x}_i'\underset{\sim}{\beta})}{\sqrt{(1-\lambda_i)\delta^2}}\}/\sqrt{(1-\lambda_i)\delta^2}\; d\underset{\sim}{\theta}$$

$$\times \prod_{i=1}^{\ell} \left\{ \sqrt{\lambda_i}\phi(\frac{\hat{\theta}_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}}{\sqrt{\delta^2/\lambda_i}}) \right\} \prod_{i=1}^{\ell} \left\{ (1/\sigma_i^2)^{(n_i+\alpha-1)/2+1} e^{-\{(n_i-1)S_i^2 + \alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\gamma}}\}/2\sigma_i^2} \right\}. \tag{13}$$

Note that the term, $\int_{\underset{\sim}{\theta} \in V} \prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})/\delta\}\; d\underset{\sim}{\theta}$, is not a function of the $\sigma_i^2$ and has been eliminated together with other such terms.

Now, one can integrate out the $\sigma_i^2$ from (11) to get the joint posterior density of $\Omega$,

$$\pi(\Omega \mid \underset{\sim}{\hat{\theta}}, \underset{\sim}{S}^2) \propto \pi(\underset{\sim}{\beta}, \delta^2, \underset{\sim}{\gamma}, \alpha) \prod_{i=1}^{\ell} \left\{ \frac{\Gamma(\alpha/2)}{(\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\gamma}}/2)^{\alpha/2}} \frac{\Gamma(n_i+\alpha-1)/2)}{\{((n_i-1)S_i^2 + \alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\gamma}})/2\}^{(n_i+\alpha-1)/2}} \right\}$$

$$\times \frac{1}{\int_{\underset{\sim}{\theta} \in V} \prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})/\delta\}\; d\underset{\sim}{\theta}} \int_{\underset{\sim}{\sigma}^2} \left[ \int_{\underset{\sim}{\theta} \in V} \prod_{i=1}^{\ell} \phi\{\frac{\theta_i - (\lambda_i \hat{\theta}_i + (1-\lambda_i)\underset{\sim}{x}_i'\underset{\sim}{\beta})}{\sqrt{(1-\lambda_i)\delta^2}}\}/\sqrt{(1-\lambda_i)\delta^2}\;\; d\underset{\sim}{\theta} \right.$$

$$\left. \times \prod_{i=1}^{\ell} \left\{ \frac{1}{\sqrt{\delta^2/\lambda_i}} \phi(\frac{\hat{\theta}_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}}{\sqrt{\delta^2/\lambda_i}}) IG_{\sigma_i^2}(a_i, b_i) \right\} \right] d\underset{\sim}{\sigma}^2, \tag{14}$$

where $a_i = (n_i + \alpha - 1)/2$ and $b_i = \{(n_i-1)S_i^2 + \alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\gamma}}\}/2$ and $IG_c(a,b)$ is the inverse gamma density, which is given by $f(c) = b^a(\frac{1}{c})^{a+1}e^{-b/c}/\Gamma(a), c > 0$.

### 3.2 Computation for the Gamma Regression Model

Our strategy is to draw samples from the joint posterior density of $\Omega$ in (14). This is a difficult task, but once this is accomplished, we can use the multiplication rule to draw samples of the $\sigma_i^2$ from (13) and then the $\theta_i$ from (12). This strategy is useful if there are a large number of counties; the state of Texas has 254 counties. We draw the $\theta_i$ in the same manner as described in Section 2. It is more difficult to draw samples of $\sigma_i^2$. We describe how to draw samples from $\Omega$ in (14). The basic strategy has two key steps.

First, we fit the double shrinkage model without the inequality constraints and the benchmarking. This gives an approximate sample of size $M = 1000$ iterates from the posterior density of $\Omega$ that we obtained using a Metropolis sampler. The details of this first step are given in Appendix A.

Second, we convert this approximate sample to a sample from the posterior density with the inequality constraint and the benchmarking. We use the $M$ iterates from the first step to construct a multivariate Student's t density for $(\underset{\sim}{\beta}, \log(\delta^2), \gamma, \log(\alpha))$. At each of the iterate obtained from the first step, we run a Metropolis sampler with the multivariate Student's t density 100 times and picked the last one. In this divide-and-conquer manner, we minimize the chance of the Metropolis sampler getting stuck. We want the Metropolis sampler to move from the starting value at least once; no other monitoring is necessary; if it does not move at least once, we discard this run. It is good that this procedure gives a sample of $M$ independent iterates of $\Omega$. However, this step is time-consuming and for the current simulated data it took roughly sixteen hours.

Now, we describe how to use the accept-reject algorithm to draw samples of $\sigma_i^2$. We can rewrite (13) as

$$\pi(\underset{\sim}{\sigma}^2 \mid \Omega, \underset{\sim}{\hat{\theta}}, \underset{\sim}{S}^2) \propto \int_{\underset{\sim}{\theta} \in \tilde{V}} \prod_{i=1}^{\ell} \phi\{\frac{\theta_i - (\lambda_i \hat{\theta}_i + (1-\lambda_i)\underset{\sim}{x}_i'\underset{\sim}{\beta})}{\sqrt{(1-\lambda_i)\delta^2}}\}/\sqrt{(1-\lambda_i)\delta^2} d\underset{\sim}{\theta}$$

$$\times \prod_{i=1}^{\ell} \left\{ \sqrt{\lambda_i}\phi(\frac{\hat{\theta}_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}}{\sqrt{\delta^2/\lambda_i}}) \right\} \int_{\underset{\sim}{\theta} \in \tilde{V}} I(\underset{\sim}{\theta} \in V) \frac{\prod_{i=1}^{\ell} \phi\{\frac{\theta_i - (\lambda_i\hat{\theta}_i + (1-\lambda_i)\underset{\sim}{x}_i'\underset{\sim}{\beta})}{\sqrt{(1-\lambda_i)\delta^2}}\}/\sqrt{(1-\lambda_i)\delta^2}}{\int_{\underset{\sim}{\theta} \in \tilde{V}} \prod_{i=1}^{\ell} \phi\{\frac{\theta_i - (\lambda_i\hat{\theta}_i + (1-\lambda_i)\underset{\sim}{x}_i'\underset{\sim}{\beta})}{\sqrt{(1-\lambda_i)\delta^2}})\}/\sqrt{(1-\lambda_i)\delta^2} d\underset{\sim}{\theta}} \, d\underset{\sim}{\theta}$$

$$\times \prod_{i=1}^{\ell} \left\{ (1/\sigma_i^2)^{(n_i+\alpha-1)/2+1} e^{-\{(n_i-1)S_i^2 + \alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\gamma}}\}/2\sigma_i^2} \right\}, \tag{15}$$

where $\tilde{V} \supseteq V$ and $\tilde{V}$ is a larger rectangular set.

Note that the first and third terms in (15) are probabilities. It is also true that the second term in (15) is a probability because $\prod_{i=1}^{\ell} \left\{ \sqrt{\lambda_i}\phi(\frac{\hat{\theta}_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}}{\sqrt{\delta^2/\lambda_i}}) \right\} \leq \{\frac{1}{\sqrt{2\pi}}\}^{\ell}$. Therefore, we can use an accept-reject sampler to draw the $\sigma_i^2$.

Note that, by construction, the first term in (15) is a product over $i = 1, \ldots, \ell$. This is also true for the second term. So if we ignore the third term, we can independently draw $\sigma_i^2 \overset{ind}{\sim} IG(a_i, b_i), i =$

$1, \ldots, \ell$ (unrestricted distributions) and take it with probability,

$$\int_{\underset{\sim}{\theta} \in \tilde{V}} \phi\{\frac{\theta_i - (\lambda_i \hat{\theta}_i + (1-\lambda_i)\underset{\sim}{x}_i'\underset{\sim}{\beta})}{\sqrt{(1-\lambda_i)\delta^2}}\}/\sqrt{(1-\lambda_i)\delta^2} d\underset{\sim}{\theta} \times \left\{ \sqrt{\lambda_i}\phi(\frac{\hat{\theta}_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}}{\sqrt{\delta^2/\lambda_i}}) \right\}$$

to complete the accept-reject algorithm. It is possible that there are several rejections before an acceptance, but this rarely happens. If there are 25 rejections, we simply draw the $\sigma_i^2$ from their unrestricted distributions, $\sigma_i^2 \overset{ind}{\sim} IG(a_i, b_i), i = 1, \ldots, \ell$.

The remaining question then is how to calculate

$$C = \int_{\underset{\sim}{\theta} \in \tilde{V}} I(\underset{\sim}{\theta} \in V) \frac{\prod_{i=1}^{\ell} \phi\{\frac{\theta_i - (\lambda_i \hat{\theta}_i + (1-\lambda_i)\underset{\sim}{x}_i'\underset{\sim}{\beta})}{\sqrt{(1-\lambda_i)\delta^2}}\}/\sqrt{(1-\lambda_i)\delta^2}}{\int_{\underset{\sim}{\theta} \in \tilde{V}} \prod_{i=1}^{\ell} \phi\{\frac{\theta_i - (\lambda_i \hat{\theta}_i + (1-\lambda_i)\underset{\sim}{x}_i'\underset{\sim}{\beta})}{\sqrt{(1-\lambda_i)\delta^2}}\}/\sqrt{(1-\lambda_i)\delta^2} d\underset{\sim}{\theta}} \, d\underset{\sim}{\theta}.$$

A Monte Carlo estimator of $C$ is

$$\hat{C} = \frac{1}{M} \sum_{h=1}^{M} I(\underset{\sim}{\theta}^{(h)} \in V),$$

where $\theta_i^{(h)} \overset{ind}{\sim} \text{Normal}\{\lambda_i \hat{\theta}_i + (1-\lambda_i)\underset{\sim}{x}_i'\underset{\sim}{\beta}, (1-\lambda_i)\delta^2\}, c_i < \theta_i^{(h)} < \infty, h = 1, \ldots, M = 1000, i = 1, \ldots, \ell$.

However, the term, $\frac{1}{M}\sum_{h=1}^{M} I(\underset{\sim}{\theta}^{(h)} \in V)$, is difficult to incorporate into the accept-reject sampler. We have overcome the difficulty in the following manner. We have computed $\hat{C}$ and found that more than 60% of the $\hat{C}$ leads to acceptance of all the $\sigma_i^2, i = 1, \ldots, \ell$. When the $\sigma_i^2$ are not accepted, we draw samples from their unrestricted distributions, $\sigma_i^2 \overset{ind}{\sim} IG(a_i, b_i), i = 1, \ldots, \ell$.

# 4 SIMULATION STUDY

Both NASS survey data and USDA administrative acreage data are subject to confidentiality protections, therefore, we describe a means of simulating data with similarity to Illinois corn crop data that have been used extensively in recent NASS studies on crop county estimates and use it to show the key features of our benchmarking procedure with inequality constraints. As a practical matter, participation in farm support programs can vary by crop and by state. Some of the survey estimates may already satisfy the lower bound constraint, i.e., some $\hat{\theta}_i > c_i$, so that the lower bound

constraints imposed on model estimates for these areas may be loose or non-binding restrictions in those counties. However, in states with high rates of enrollment in farm support programs, like the corn crop in Illinois, administrative totals may capture large parts of the population, so that direct estimates, subject to sampling error, fall below to the administrative totals in many counties. The model estimates of the counties must be constrained by the lower bounds and the benchmarking target as well.

In Section 4.1, we describe the simulation plan, where we describe several simulated data sets. In Section 4.2, we present results under the single shrinkage model with the inequality constraints. In Section 4.3, we present results under the double shrinkage model for the gamma regression model and the log-linear model, again with the inequality constraints. At the same time, we have compared these models with the direct estimates (DE), the estimates from the Bayesian Fay-Herriot model (ME), without benchmarking or inequality constraints, and the Bayesian Fay-Herriot model with random benchmarking (MERB) at both the county level and at the level of agricultural statistic districts (discussed below).

### 4.1 Description of the Simulated Data Sets

Nandram, Erciulescu, and Cruze (2019) simulated a data set similar to the one in Battese, Harter and Fuller (1988); see also Toto and Nandram (2010) and Nandram, Toto and Choi (2011). These data are on planted acres of corn and soybeans for 37 segments with 12 counties in the state of Iowa and there are two covariates. (Like Illinois, Iowa is a large corn producing state in the United States.) By simulating from these data, we can create a data set with as many areas we please. In particular Illinois has $\ell = 102$ counties grouped in 9 smaller-than-state regions called Agricultural Statistics Districts (ASDs). The data are processed to obtain the survey estimates and standard errors. In our simulated data, based on the actual sizes of the ASDs, we have taken the first set of counties to be in the first ASD, the second set to be in the second ASD and so on so that the first 12 counties correspond to the first ASD, the next 11 correspond to the second, and the remaining ASDs have 9, 11, 7, 13, 15, 12, 12 counties, respectively. In the process of simulating acreage data, we also added a random effect for each ASD. The sample sizes within the counties are chosen uniformly in $(2, 74)$, a realistic range of sample sizes across the state comparable to actual Illinois corn data reported during the 2014 crop year (Erciulescu, Cruze, and Nandram

2018, 2019). Additionally, county-level coefficients of variation $CV_i$ will be simulated uniformly from within the range of $(.08, .93)$; these extremes are comparable to values reported in Erciulescu, Cruze, and Nandram (2020) in reference to the 2015 crop year. Given simulated survey estimates and coefficients of variation, computed standard errors are obtained $\hat{\sigma}_i = CV_i \times \hat{\theta}_i$. Thus, we have a data set with the survey estimates, $\hat{\theta}_i$, survey standard error, $\hat{\sigma}_i$ and sample sizes, $n_i$ for the $i^{th}$ county, $i = 1, \ldots, \ell$.

The last piece to be simulated is the data corresponding to the administrative acreage values, i.e., lower bounds, $c_i$. For simplicity, we call these the FSA values throughout the simulation. In order to reflect the relationship between the FSA values and the survey estimates for Illinois, we assume the following equation holds,

$$c_i = \hat{\theta}_i + U_i \times \hat{\theta}_i, i = 1, \ldots, \ell,$$

where $U_i \overset{iid}{\sim}$ Uniform$(-s, s)$ and $s$ is taken to be a suitable value (e.g., $s = .10$). However, the key problem is how to set the benchmarking target. In the real problem, we will know the target, but the target has to be larger than the sum of the lower bounds. Therefore, it is sensible to take the target to be $a = c/d$, where $c \equiv \sum_{i=1}^{\ell} c_i$ and specify $0 < d < 1$. The completeness of the administrative data relative to the state total can vary by state and crop, but in Illinois, this value will often be close to 1.

### 4.2 Results under the Single Shrinkage Model

In applying the methodology for an inequality-constrained model with fixed variances developed in Section 2, we specify a plausible value of $d = .99$ indicating the simulated administrative data embody 99% of the state-level planted area total for corn in Illinois. In this first instance, we restrict the range of coefficients of variation to $(0.05, 0.25)$. Figure 1 shows the simulated survey estimates of $\hat{\theta}_i$ versus the FSA values $c_i$ (top panel) and the posterior mean of $\theta_i$ versus the FSA values $c_i$ under the Bayesian Fay-Herriot model with inequality constraint and benchmarking, not including double shrinkage (call this model MFSA-NDS). In the top panel, we can see many points are above or below the $45^o$ straight line through the origin. (This resembles a realistic pattern shown in Figure 4 of Erciulescu, Cruze, and Nandram (2020), as applied to the 2015 Illinois corn crop.) Where the

survey estimates for many counties are below their corresponding FSA values, all points in the bottom are immediately above the $45^o$ straight line through the origin, indicating that all MFSA-NDS estimates are no smaller than their corresponding FSA values. Moreover, the sum of the 102 MFSA is equal to the state total, satisfying the benchmarking requirement by raking to the state target.

In Table 1, we present results for Illinois simulated data. We compare the results with our new model that incorporates the inequality constraints (FSA values are lower bounds of the model estimates). Specifically, we compare estimates from DE, ME and MERB and the single shrinkage Bayesian Fay-Herriot model with inequality constraint and benchmarking (MFSA-NDS).

The minimum, median and maximum posterior coefficients of variation (expressed as percents, %) are smaller than the other two models (ME, MERB), even more so for the direct estimates (DE). Of course, as expected, the coefficients of variation for the ASDs are smaller than those for the counties; there is one exception (5.13 versus 5.31 in Table 1). We note that, as expected, the coefficients of variation are in decreasing order (DE, ME, MERB, MFSA), and modeling appears beneficial, but more importantly we can accommodate the FSA values in our model (MFSA) and provide much smaller coefficients of variation.

Table 1: Coefficients of Variation (%) for Illinois simulated data for 102 counties and 9 Agricultural Statistical Districts, fixed variances

| Level | Statistic | DE | ME | MERB | MFSA-NDS |
|---|---|---|---|---|---|
| County | min | 5.13 | 4.76 | 4.79 | 0.57 |
| | median | 15.57 | 10.67 | 10.58 | 0.97 |
| | max | 24.93 | 15.80 | 15.34 | 5.22 |
| ASD | min | 5.31 | 2.54 | 2.39 | 0.24 |
| | median | 10.60 | 3.25 | 3.01 | 0.30 |
| | max | 14.81 | 3.92 | 3.51 | 0.39 |

NOTE: MFSA is the new benchmarking model with FSA values as lower bounds for the model estimates, CV $(.05 - .25)$ and $d = .99$.

### 4.3 Results under the Double Shrinkage Model

Fitting the double shrinkage model with the inequality constraints of Section 3 and denoting these estimates as MFSA-DS, we fit the model to the data already generated for Section 4.2. That

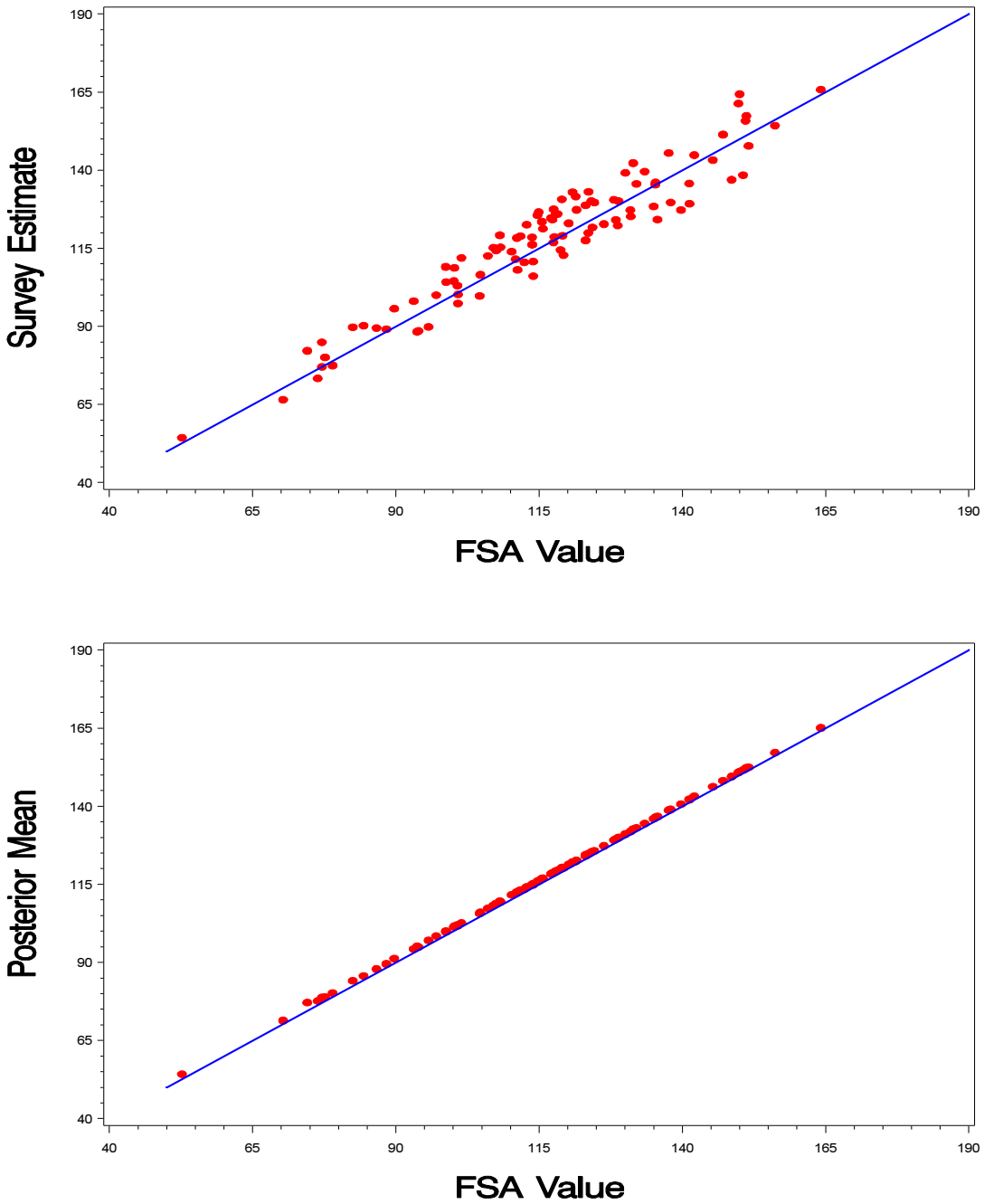


Figure 1: Plots of survey estimates (top panel) and posterior means (bottom panel) under MFSA-NDS for $\theta$ versus FSA values for Illinois and the simulated data, not double shrinkage, CV $(.05-.25)$ and $d = .99$.

is, data for which the simulated $CV_i \in (0.05, 0.25)$ and $d = .99$. Summaries of the coefficients in variation for the MFSA-DS are given in Table 2, with the first four columns duplicated from Table 1. We notice a small difference between the double shrinkage model and the single shrinkage model. Over counties the maximum $CV$ under the double shrinkage model is a bit smaller than the one under the MFSA-NDS models, 3.58% versus 5.22% for the fixed-variances case, but over ASDs (aggregates of counties within) there are smaller differences between the two approaches.

The top panel in Figure 2 once again plots the survey estimates versus FSA values (identical to top panel, Figure 1), and the lower panel is a plot of the posterior means versus the FSA values under the double shrinkage model with benchmarking and inequality constraints. The lower panel of Figure 2 is only slightly different from that of Figure 1, in part because the value $d = 0.99$ implies that there is little slack between the state target and the total of administrative data summed over all counties in the state.

Table 2: Coefficients of Variation (%) for Illinois simulated data for 102 counties and 9 Agricultural Statistical Districts, double shrinkage, gamma model

| Level | Statistic | DE | ME | MERB | MFSA-NDS | MFSA-DS |
|---|---|---|---|---|---|---|
| County | min | 5.13 | 4.76 | 4.79 | 0.57 | 0.55 |
| | median | 15.57 | 10.57 | 10.58 | 0.97 | 1.01 |
| | max | 24.93 | 15.80 | 15.34 | 5.22 | 3.58 |
| ASD | min | 5.31 | 2.54 | 2.39 | 0.24 | 0.26 |
| | median | 10.60 | 3.25 | 3.01 | 0.30 | 0.34 |
| | max | 14.81 | 3.92 | 3.51 | 0.42 | 0.41 |

NOTE: MFSA is the new benchmarking model with FSA values as lower bounds for the model estimates. MFSA-DS refers to the double shrinkage model with benchmarking and inequality constraint, CV $(.05 - .25)$ and $d = .99$.

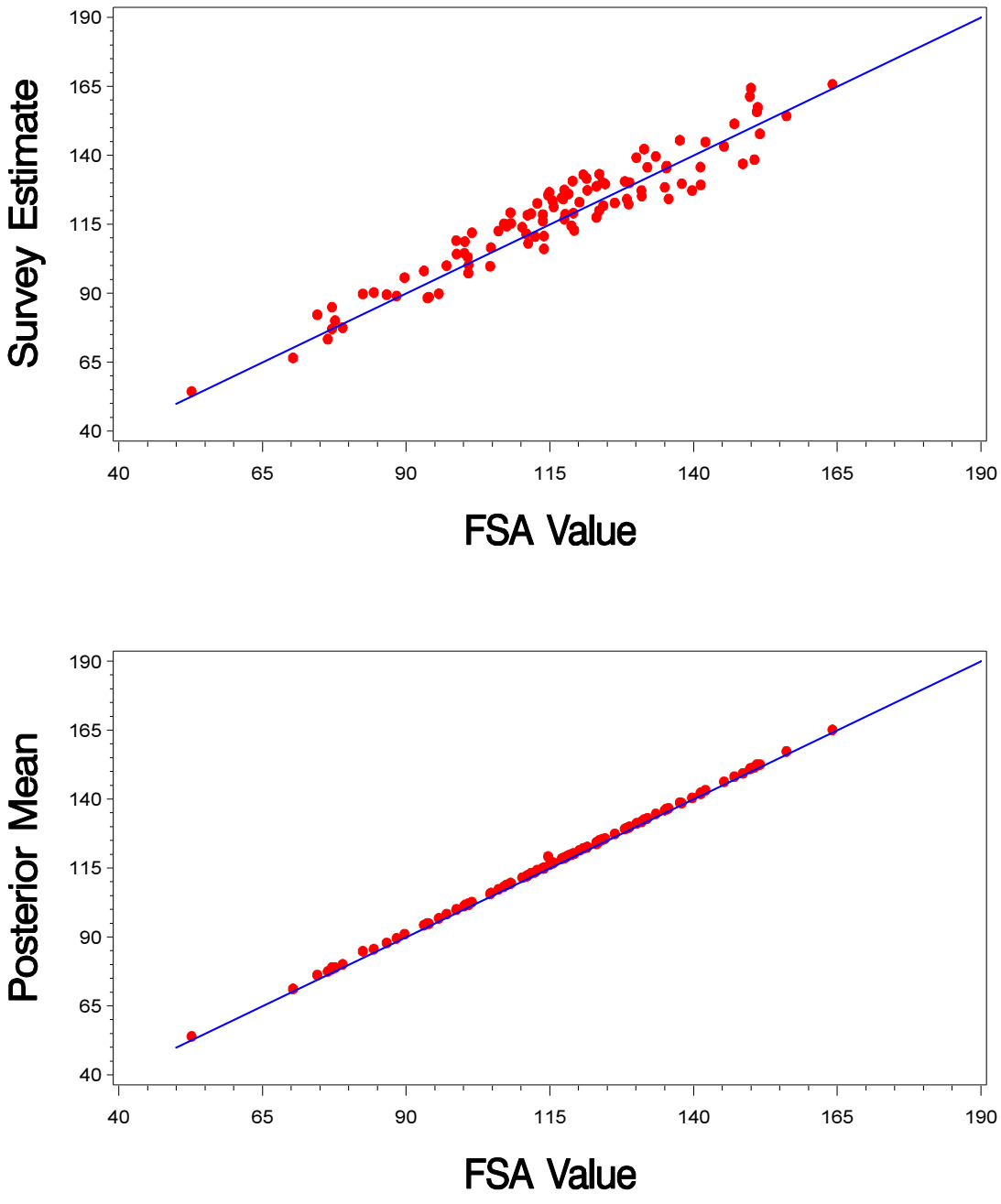


Figure 2: Plots of survey estimates (top panel) and posterior means (bottom panel) under MFSA-DS for $\theta$ versus FSA values for Illinois and the simulated data, double shrinkage, gamma regression, cv $(.05 - .25)$ and $d = .99$.

For the purposes of demonstrating the log-linear model, a second data set with slightly different features has been generated. Namely, we specify lower coverage of the FSA values ($d = 0.95$) and allow a higher range of values of survey coefficients of variation, $(0.08, 0.93)$, comparable to the actual survey coefficients of variation observed during the 2015 crop year. We present summaries of the CVs in Table 3. Again we notice a small difference between the log-linear double shrinkage model and the single shrinkage model at the ASD level. Differences in coefficients of variation at the county level are minimal for the lower half of all counties, but the maximum county CV obtained from the double shrinkage model (23.94%) is substantially smaller than the maximum CV obtained under the single shrinkage model (fixed-variances case) (44.92%).

In its upper panel, Figure 3 depicts the new simulated survey estimates versus their corresponding FSA values, while the lower panel shows the posterior means of the log-linear MFSA-DS model versus the corresponding FSA values. In contrast to the $d = .99$ data set of the previous sections, the present $d = .95$ data set represents a looser lower-bound constraint. Accordingly, the resulting county acreage estimates, which also sum to the state total, are all visibly above the $45^o$ line. For comparison, the MFSA-DS estimates obtained under gamma regression are plotted in the lower panel of Figure 4. The two approaches to double shrinkage yield similar (not identical) point estimates given the same state target and administrative lower bound constraints.

In contrast to the computationally expensive gamma regression which required in excess of 16 hours of run time, results of the log-linear model were obtained in a matter of minutes, and additional opportunities to speed up the process may be possible through approximate Bayesian computation described in Appendix B. It is worth noting that all samplers were coded in Fortran 90 on a server with Intel Xeon E5-2690 2.90GHz processor with eight cores.

# 5 CONCLUSIONS

Beginning with the 2020 crop year, NASS successfully converted its county-estimates data product into a system model-based estimates of planted area, harvested area, total production, and yield per harvested acre. The official estimates for 13 different commodity crops grown nationwide now include a benchmarking of county estimates to predetermined state targets, and lower bound constraints on planted area. Motivated by the needs of the NASS crop estimation program to produce

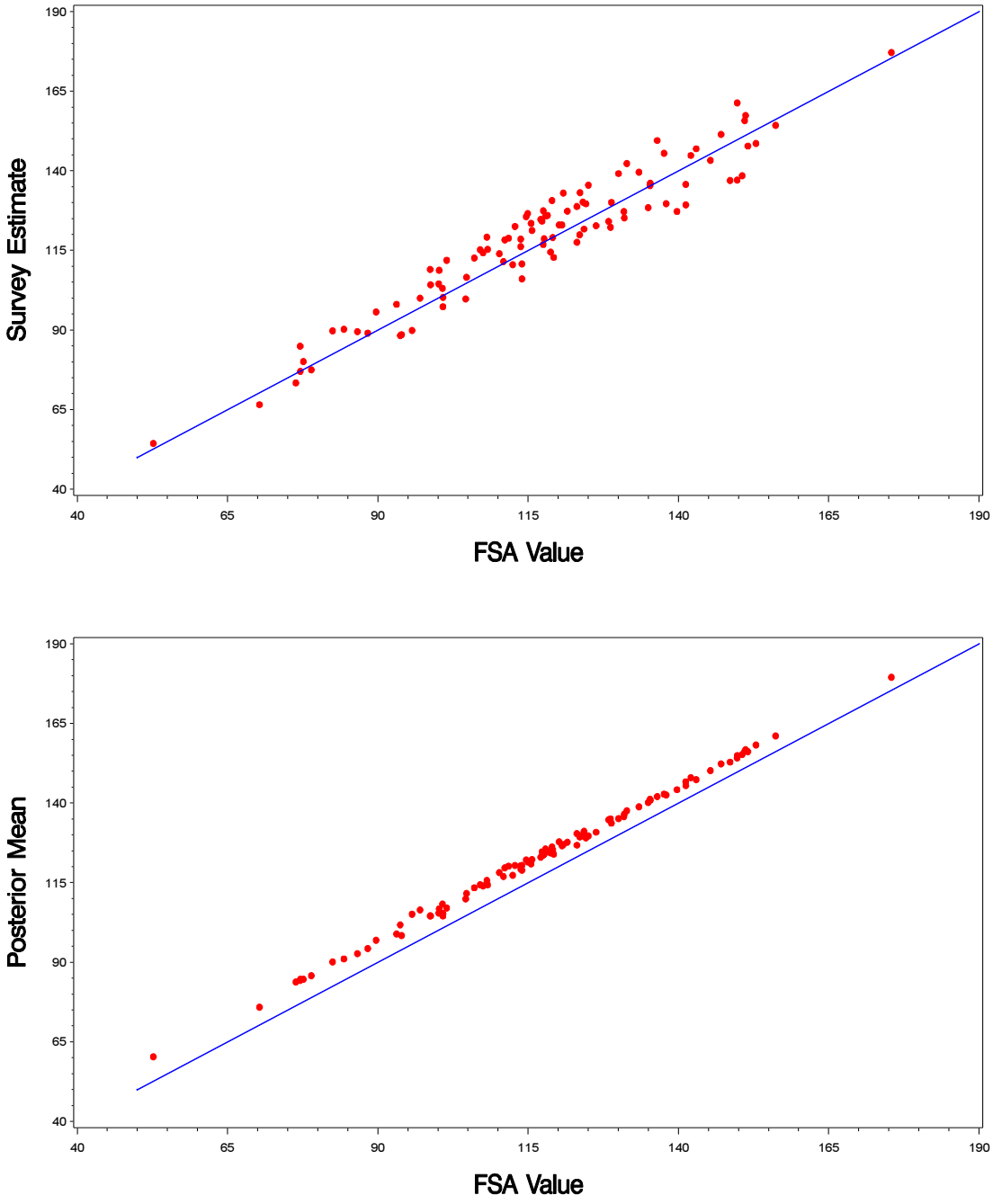


Figure 3: Plots of survey estimates (top panel) and posterior means (bottom panel) under MFSA for $\theta$ versus FSA values for Illinois and the simulated data, double shrinkage, **log-linear model**; CV $(.08 - .93)$ and $d = .95$

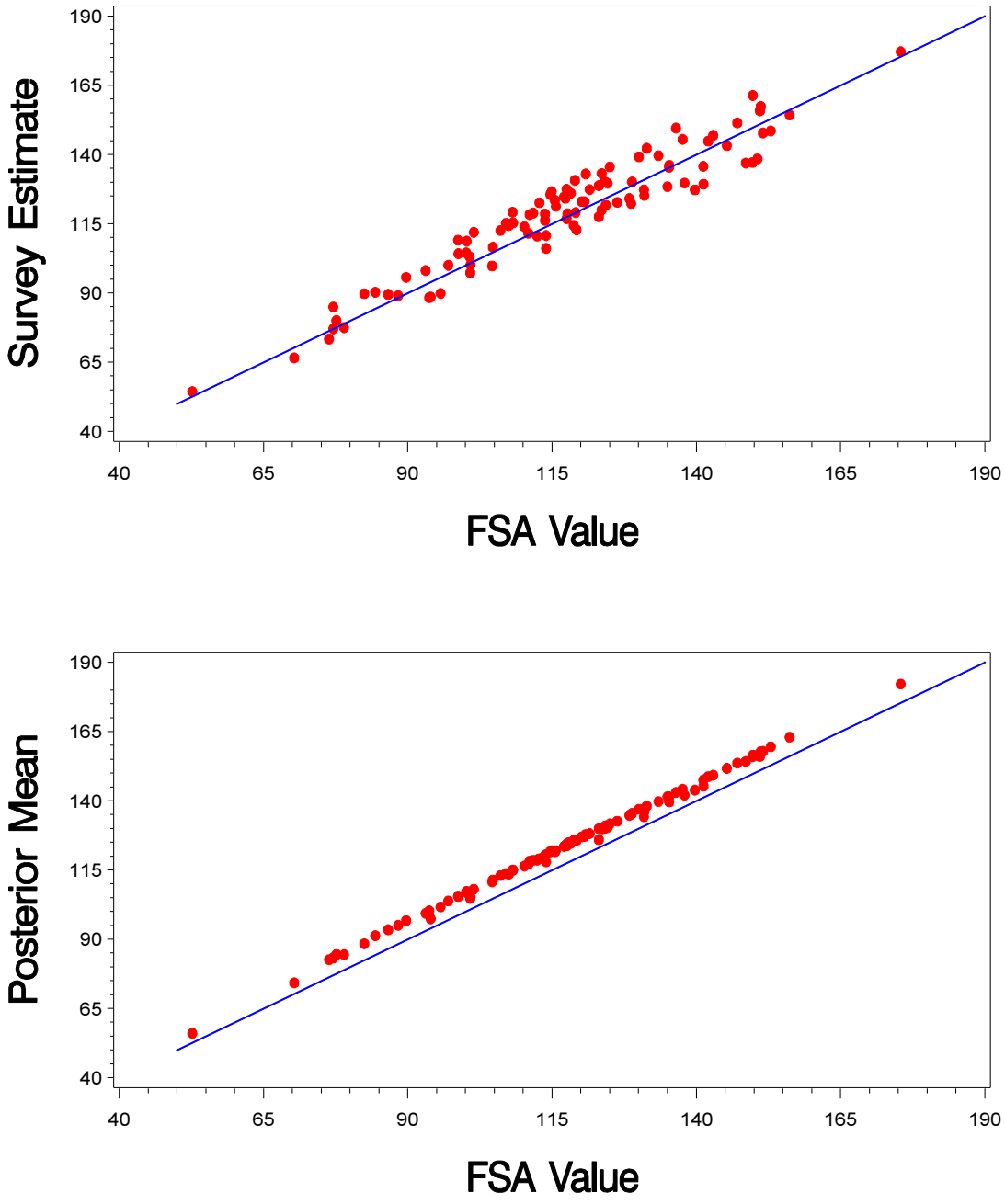


Figure 4: Plots of survey estimates (top panel) and posterior means (bottom panel) under MFSA-DS for $\theta$ versus FSA values for Illinois and the simulated data, double shrinkage, **gamma regression**, CV $(.08 - .93)$ and $d = .95$

Table 3: Coefficients of Variation (%) for Illinois simulated data for 102 counties and 9 Agricultural Statistical Districts, double shrinkage, log-linear model

| Level | Statistic | DE | ME | MERB | MFSA-NDS | MFSA-DS |
|---|---|---|---|---|---|---|
| County | min | 8.57 | 7.73 | 7.90 | 2.57 | 2.54 |
| | median | 52.90 | 17.83 | 17.16 | 4.56 | 4.92 |
| | max | 92.70 | 24.25 | 24.82 | 44.92 | 23.94 |
| ASD | min | 18.90 | 5.11 | 3.78 | 1.17 | 1.11 |
| | median | 37.70 | 6.15 | 4.71 | 1.83 | 1.43 |
| | max | 52.10 | 7.19 | 5.81 | 2.65 | 1.63 |

NOTE: MFSA is the new benchmarking model with FSA values as lower bounds for the model estimates. MFSA-DS refers to the double shrinkage model with benchmarking and inequality constraint, CV $(.08 - .93)$ and $d = .95$.

coherent published tables across all parameters and with respect available administrative data, we have shown how to incorporate the area-specific inequality constraints *and benchmarking* into the Fay-Herriot model. Single shrinkage model and double shrinkage models are available. Because there are difficulties in performing full Metropolis samplers, we overcame these computational difficulties by making additional reasonable approximations in the double shrinkage model.

It is possible to extend the hierarchical Bayesian model so that all the constraints are actually included in it. That is, $\underset{\sim}{\theta}$ is in $V = \{\underset{\sim}{\theta} : c_i \leq \theta_i \sum_{i=1}^{n} \theta_i = a\}$, where $a$ is the benchmarking target and $c_1, \ldots, c_n$ are the FSA values. So that the hierarchical Bayesian model (i.e., extended version of the Bayesian Fay-Herriot model) has $\underset{\sim}{\theta} \in V$. We have attempted to do so for the simplest model, the Bayesian Fay-Herriot model, but the problem is extremely difficult. It requires the computation of orthant probabilities (e.g., Ridgway 2016, Geweke 1991, Genz 1992) at each step of a Markov chain Monte Carlo sampler. There are no such problems mentioned in Rao and Molina (2015), although they have used the raking procedure for benchmarking only, not the inequality constraints, where the $\theta_i > c_i$, the FSA problem.

Nevertheless, incorporating the total constraint into the hierarchical Bayesian model will be beneficial because it will help protect against model failure so prominent in small area estimation, and one needs to be careful with this. Toto and Nandram (2010), Nandram and Sayit (2011) and Nandram, Toto and Choi (2010), Nandram, Erciulescu and Cruze (2019) and Janicki and Vesper (2017) were able to incorporate a much simpler constraint (i.e., $\sum_{i=1}^{n} \theta_i = a$) in a complete Bayesian

analysis. But as is evident, it is much more difficult to incorporate the constraint $\underset{\sim}{\theta} \in V$, and it is a problem we would like to solve in the future. We can add random effects on both means and variances to accommodate sub-areas (counties within ASDs). However, the computations are difficult and approximations beyond those based on Markov chain Monte Carlo methods need to be considered. Currently, we are doing research in this area.

In Appendix C, we have comments on generalizion. It is possible to avoid the inequality constraint using a logarithmic transformation, but this method looses generality or it makes unnecessary approximation. Our solution remains strong for both the single shrinkage model and the double shrinkage model.

# APPENDIX A Double-Shrinkage Model Fitting–Gamma Regression

Dropping the inequality constraint of the double shrinkage model (see 11), the joint posterior density is

$$\pi(\underset{\sim}{\theta}, \underset{\sim}{\sigma}^2, \Omega \mid \underset{\sim}{\hat{\theta}}, \underset{\sim}{S}^2) \propto \pi(\underset{\sim}{\beta}, \delta^2, \underset{\sim}{\gamma}, \alpha) \prod_{i=1}^{\ell} \left\{ (\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\gamma}}/2)^{\alpha/2} (1/\sigma_i^2)^{\alpha/2+1} e^{-(\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\gamma}}/2\sigma_i^2)} / \Gamma(\alpha/2) \right\}$$

$$\times \prod_{i=1}^{\ell} \left\{ \frac{1}{\sqrt{(1-\lambda_i)\delta^2}} \phi \left( \frac{\theta_i - (\lambda_i \hat{\theta}_i + (1-\lambda_i)\underset{\sim}{x}_i'\underset{\sim}{\beta})}{\sqrt{(1-\lambda_i)\delta^2}} \right) \frac{1}{\sqrt{\delta^2/\lambda_i}} \phi\left(\frac{\hat{\theta}_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}}{\sqrt{\delta^2/\lambda_i}}\right) \right\}$$

$$\times \prod_{i=1}^{\ell} \left\{ [(n_i-1)/2\sigma_i^2]^{(n_i-1)s_i^2/2} e^{-(n_i-1)/2\sigma_i^2} \right\}, \tag{A.1}$$

where $\lambda_i = \delta^2/(\delta^2+\sigma_i^2), i = 1, \ldots, \ell$. Conditional on $\Omega, \underset{\sim}{\hat{\theta}}, \underset{\sim}{S}^2$, it is clear that $(\theta_i, \sigma_i^2)$ are independent over $i = 1, \ldots, \ell$. This is the key difference between the double-shrinkage model with and without the inequality constraints.

Our strategy is to first sample the posterior density $\pi(\Omega \mid \underset{\sim}{\hat{\theta}}, \underset{\sim}{S}^2)$. Once this is done, we draw samples from the joint conditional posterior density of $\pi(\underset{\sim}{\sigma}^2 \mid \Omega, \underset{\sim}{\hat{\theta}}, \underset{\sim}{S}^2)$. Then, finally we obtain the required samples from $\pi(\underset{\sim}{\theta} \mid \underset{\sim}{\sigma}^2, \Omega, \underset{\sim}{\hat{\theta}}, \underset{\sim}{S}^2)$. Thus, after draws are obtained for $\Omega$, we use the

multiplication rule to get the $\sigma_i^2$ and $\theta_i$ (i.e., $\Omega$, the $\sigma_i^2$ and $\theta_i$ are drawn simultaneously).

It follows from (A.1) that the conditional on $\underset{\sim}{\sigma}^2, \Omega, \underset{\sim}{\hat{\theta}}, \underset{\sim}{S}^2$, the $\theta_i$ are independent and

$$\theta_i \mid \underset{\sim}{\sigma}^2, \Omega, \underset{\sim}{\hat{\theta}}, \underset{\sim}{S}^2 \overset{ind}{\sim} \text{Normal}\{\lambda_i\hat{\theta}_i + (1-\lambda_i)\underset{\sim}{x}_i'\underset{\sim}{\beta}, (1-\lambda_i)\delta^2\}, i = 1, \ldots, \ell. \tag{A.2}$$

Conditional on $\Omega, \underset{\sim}{\hat{\theta}}, \underset{\sim}{S}^2$, the $\sigma_i^2$ are independent. Therefore, integrating out the $\theta_i$ from (A.1), we have the conditional posterior density of $\sigma_i^2$ is

$$\pi(\sigma_i^2 \mid \Omega, \underset{\sim}{\hat{\theta}}, \underset{\sim}{S}^2) \propto \sqrt{\lambda_i}\phi(\frac{\hat{\theta}_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}}{\sqrt{\delta^2/\lambda_i}})[(1/\sigma_i^2)^{(n_i+\alpha-1)/2+1} e^{-\{(n_i-1)S_i^2 + \alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\gamma}}\}/2\sigma_i^2}], \tag{A.3}$$

$i = 1, \ldots, \ell$. Note unnecessary constants are dropped (e.g., parameters conditioned on).

Now, one can integrate out the $\theta_i$ and $\sigma_i^2$ from (A.1) to get the joint posterior density of $\Omega$,

$$\pi(\Omega \mid \underset{\sim}{\hat{\theta}}, \underset{\sim}{S}^2) \propto \pi(\underset{\sim}{\beta}, \delta^2, \underset{\sim}{\gamma}, \alpha) \prod_{i=1}^{\ell} \left\{ \frac{\Gamma(\alpha/2)}{(\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\gamma}}/2)^{\alpha/2}} \frac{\Gamma(n_i+\alpha-1)/2)}{\{((n_i-1)S_i^2 + \alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\gamma}})/2\}^{(n_i+\alpha-1)/2}} \right\}$$

$$\times \prod_{i=1}^{\ell} \left\{ \int_0^{\infty} \frac{1}{\sqrt{\delta^2/\lambda_i}} \phi(\frac{\hat{\theta}_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}}{\sqrt{\delta^2/\lambda_i}}) IG_{\sigma_i^2}(a_i, b_i) d\sigma_i^2 \right\}, \tag{A.4}$$

where $a_i = (n_i + \alpha - 1)/2$ and $b_i = \{(n_i-1)S_i^2 + \alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\gamma}}\}/2$. Here, $IG_x(a, b)$ is the inverse gamma density and is given by $f(x) = b^a(\frac{1}{x})^{a+1}e^{-b/x}/\Gamma(a), x > 0$.

It is easy to sample the $\sigma_i^2$ in (A.3) using the accept-reject sampler; simply draw $\sigma_i^2 \mid \Omega, \underset{\sim}{\hat{\theta}}, \underset{\sim}{S}^2 \sim IG(a_i, b_i)$ and take it with probability $\sqrt{\lambda_i}\phi(\frac{\hat{\theta}_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}}{\sqrt{\delta^2/\lambda_i}})$. Then, clearly the $\theta_i$ are easy to draw from (A.2). The main problem now is how to sample the joint posterior density of $\Omega$ in (A.4). We will use the Metropolis sampler to do so.

Once we obtain a sample from (A.4), we convert it to a sample from (14), our main objective. This is accommodated by another Metropolis sampler that we execute in a novel manner. We prefer to use proposal densities that will provide independent chains. This is obtained by taking draws from a multivariate Student's t density (to be constructed). We will not use a long run because with a Metropolis sampler, the chain tends to get stuck a long time, introducing long-range dependence to the sample, thereby giving poor mixing that is inefficient. Instead we run several chains, say

$M = 1000$ chains. Each chain is run with a random start from an approximate density for 100 iterates, and the last one is taken. Only minor monitoring is needed to ensure reasonable jumping rates. If the chain does not move from the initial random start, it is not used in the final sample. In the end, we get a random sample of $M$ iterates from the required density in (A.4).

We describe how to obtain samples from the posterior density of $\Omega = (\underset{\sim}{\beta}, \delta^2, \underset{\sim}{\gamma}, \alpha)$. There are three steps. The first step obtains a sample of $M$ starting values, the second step is to obtain a proposal density for Metropolis sampler at each of the starting values, and the third step is to make a short run of 100 iterates of each of the Metropolis samplers in second step.

First, we integrate out the $\theta_i$ and we replace the $\sigma_i^2$ by $S_i^2, i = 1, \ldots, \ell$. Given $\underset{\sim}{\hat{\theta}}, \underset{\sim}{S}^2$, then $(\underset{\sim}{\beta}, \delta^2)$ and $(\underset{\sim}{\gamma}, \alpha)$ are independent; so they can be sampled separately to get $M = 1000$ independent starts. We have obtained these $M$ starts using simple approximations.

Second, at each start, we run a Gibbs sampler to get $\sigma_i^2, i = 1, \ldots, \ell$, and $\Omega$. This is done by drawing the $\sigma_i^2$ from their exact conditional posterior densities using rejection sampling. Then, given $\underset{\sim}{\sigma}^2$, $\underset{\sim}{S}^2$, $(\underset{\sim}{\beta}, \delta^2)$ and $(\underset{\sim}{\gamma}, \alpha)$ are again independent, and draws from their respective joint posterior densities are taken in a similar manner. It is worth noting that given $\delta^2$, the distribution of $\underset{\sim}{\beta}$ is multivariate normal and $\underset{\sim}{\beta}$ can be integrated out to get the conditional posterior density of $\delta^2$ that can be sampled using a grid. However, this is not the case for $(\underset{\sim}{\gamma}, \alpha)$ because the conditional posterior density of $\underset{\sim}{\gamma}$ given $\alpha$ is nonstandard (i.e., not multivariate normal). Thus, we approximate the posterior density of $\underset{\sim}{\gamma}$ using a multivariate normal density, and with this approximation, sampling of $(\underset{\sim}{\gamma}, \alpha)$ takes place in the same manner as for $(\underset{\sim}{\beta}, \delta^2)$.

Third, we run the second step 1100 times with a "burn-in" of 100 runs and we use the $M = 1000$ samples to construct a multivariate Student's $t$ density for $\Omega_a = (\underset{\sim}{\beta}, \log(\delta^2), \underset{\sim}{\gamma}, \log(\alpha))$, which we use as a proposal density in a Metropolis sampler to sample the exact posterior density. This is performed 100 times and the last iterate is selected. Each random start contributes to the sample of $M = 1000$ iterates of $\Omega_a$ or $\Omega$ from the posterior density under the double shrinkage model without the inequality constraint and the benchmarking.

To complete the entire procedure, for each $\Omega_a$, we sample $\sigma_i^2$ from their conditional posterior densities using rejection sampling to access the posterior densities more efficiently. Then, more importantly, the $\theta_i$ are drawn from their conditional posterior densities (normal is this case). The entire procedure took roughly four hours, and the jumping rates are mostly larger than 5%.

# APPENDIX B Doubel Shrinkage Model Fitting–Log-Linear Model

We describe the double shrinkage log-linear model and show how to fit. The main purpose is to show that there are additional gains in computational speed using approximate Bayesian computation.

Our model is similar to the one in Section 3, where assuming that $\hat{\theta}_i$ and $S_i^2$ are pairwise independent,

$$\hat{\theta}_i \mid \theta_i, \sigma_i^2 \overset{ind}{\sim} \text{Normal}(\theta_i, \sigma_i^2), i = 1, \ldots, \ell,$$

$$\frac{(n_i - 1)S_i^2}{\sigma_i^2} \mid \sigma_i^2 \overset{ind}{\sim} \text{Gamma}(\frac{n_i - 1}{2}, \frac{1}{2}), i = 1, \ldots, \ell.$$

However, a priori, we assume that

$$\theta_i \mid \underset{\sim}{\beta}_1, \delta_1^2 \overset{ind}{\sim} \text{Normal}(\underset{\sim}{x}_i'\underset{\sim}{\beta}_1, \delta_1^2), i = 1, \ldots, \ell, \underset{\sim}{\theta} \in V,$$

with the log-linear model on the $\sigma_i^2$,

$$\ln(\sigma_i^2) \mid \underset{\sim}{\beta}_2, \delta_2^2 \overset{ind}{\sim} \text{Normal}(\underset{\sim}{x}_i'\underset{\sim}{\beta}_2, \delta_2^2), i = 1, \ldots, \ell,$$

where we also assume that $\theta_i$ and $\sigma_i^2$ are pairwise independent. Note that we also have the restriction $\underset{\sim}{\theta} \in V$. Because we will use an approximate Gibbs sampler to fit the model, we assume that $\pi(\underset{\sim}{\beta}_1, \underset{\sim}{\beta}_2, \delta_1^2, \delta_2^2) \propto \frac{1}{\delta_1^2}\frac{1}{\delta_2^2}$ (i.e., posterior propriety is not an issue provided that the design matrix is full rank).

Then, letting $D = (\underset{\sim}{\hat{\theta}}, \underset{\sim}{s}^2)$, the joint posterior density of $\underset{\sim}{\theta}, \underset{\sim}{\sigma}^2, \underset{\sim}{\beta}_1, \delta_1^2, \underset{\sim}{\beta}_2, \delta_2^2$ is given by

$$\pi(\underset{\sim}{\theta}, \underset{\sim}{\sigma}^2, \underset{\sim}{\beta}_1, \delta_1^2, \underset{\sim}{\beta}_2, \delta_2^2 \mid D) \propto \prod_{i=1}^{\ell} \left\{ \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-(\hat{\theta}_i - \theta_i)^2/2\sigma_i^2} \right\} \frac{\prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}_1)/\delta_1\}}{\int_{\underset{\sim}{\theta} \in V} \prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}_1)/\delta_1\} d\underset{\sim}{\theta}}$$

$$\times \frac{1}{\delta_1^2}\frac{1}{\delta_2^2} \prod_{i=1}^{\ell} \left\{ (\frac{n_i - 1}{\sigma_i^2})^{(n_i-1)/2} e^{-(n_i-1)s_i^2/2\sigma_i^2} \frac{1}{\sqrt{2\pi\delta_2^2}} e^{-(\ln(\sigma_i^2) - \underset{\sim}{x}_i'\underset{\sim}{\beta}_2)^2/2\delta_2^2} \right\}, \underset{\sim}{\theta} \in V.$$

Our strategy in the computation is to sample the exact conditional posterior density of $\theta_i, i = 1, \ldots, \ell$, and $\sigma_i^2, i = 1, \ldots, \ell$. However, we want to replace the conditional posterior densities of $\underset{\sim}{\beta}_1, \delta_1^2$ and $\underset{\sim}{\beta}_2, \delta_2^2$ by approximate posterior densities. The main issue now is how to do this latter

task.

We consider the two simpler models for $\hat{\theta}_i$ and $s_i^2$, $i = 1, \ldots, \ell$. These are

$$\hat{\theta}_i \mid \underset{\sim}{\beta}_1, \delta_1^2 \overset{ind}{\sim} \text{Normal}(\underset{\sim}{x}_i'\underset{\sim}{\beta}_1, \delta_1^2), i = 1, \ldots, \ell, \pi(\underset{\sim}{\beta}_1, \delta_1^2) \propto 1/\delta_1^2,$$

and

$$\ln(s_i^2) \mid \underset{\sim}{\beta}_2, \delta_2^2 \overset{ind}{\sim} \text{Normal}(\underset{\sim}{x}_i'\underset{\sim}{\beta}_2, \delta_2^2), i = 1, \ldots, \ell, \pi(\underset{\sim}{\beta}_2, \delta_2^2) \propto 1/\delta_2^2.$$

Note that in the full model, we simply replace the $\theta_i$ by $\hat{\theta}_i$ and $\sigma_i^2$ by $s_i^2$. Here, the posterior densities of $(\underset{\sim}{\beta}_1, \delta_1^2)$ and $\underset{\sim}{\beta}_2, \delta_2^2)$, which are independent, have simple forms. Letting $X$ denote the $n \times p$ design matrix, then

$$\underset{\sim}{\beta}_1 \mid \underset{\sim}{\hat{\theta}}, \delta_1^2 \sim \text{Normal}\{\hat{\beta}_1, (X'X)^{-1}\delta_1^2\}, \delta_1^2 \mid \underset{\sim}{\hat{\theta}} \sim \text{IG}\left\{\frac{n-p}{2}, \frac{\sum_{i=1}^n (\hat{\theta}_i - \underset{\sim}{x}_i'\hat{\beta}_1)^2}{2}\right\},$$

where $\hat{\beta}_1 = (X'X)^{-1}X'\underset{\sim}{\hat{\theta}}$. Therefore, the posterior density of $\underset{\sim}{\beta}_1$ is a multivariate Student's $t$ density, and, in this case, it is easy to draw samples of $\underset{\sim}{\beta}_1$ and $\delta_1^2$. In addition, letting $z_i = \ln(s_i^2), i = 1, \ldots, \ell$, then

$$\underset{\sim}{\beta}_2 \mid \underset{\sim}{z}, \delta_2^2 \sim \text{Normal}\{\hat{\beta}_2, (X'X)^{-1}\delta_2^2\}, \delta_2^2 \mid \underset{\sim}{z} \sim \text{IG}\left\{\frac{n-p}{2}, \frac{\sum_{i=1}^n (z_i - \underset{\sim}{x}_i'\hat{\beta}_2)^2}{2}\right\},$$

where $\hat{\beta}_2 = (X'X)^{-1}X'\underset{\sim}{z}$. Again, the posterior density of $\underset{\sim}{\beta}_2$ is a multivariate Student's $t$ density, and it is easy to draw samples of $\underset{\sim}{\beta}_2$ and $\delta_2^2$. Our approximate Gibbs sampler runs by taking these posterior densities as the conditional posterior densities. We need to do so because the computation is difficult and time-consuming.

The joint density of $(\theta_i, \sigma_i^2), i = 1, \ldots, \ell$, is

$$\pi(\underset{\sim}{\theta}, \underset{\sim}{\sigma^2} \mid \underset{\sim}{\beta}_1, \delta_1^2, \underset{\sim}{\beta}_2, \delta_2^2, D) \propto \prod_{i=1}^{\ell} \left\{\frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-(\hat{\theta}_i - \theta_i)^2/2\sigma_i^2}\right\} \frac{\prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}_1)/\delta_1\}}{\int_{\underset{\sim}{\theta} \in V} \prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}_1)/\delta_1\} d\underset{\sim}{\theta}}$$

$$\times \prod_{i=1}^{\ell} \left\{(\frac{n_i - 1}{\sigma_i^2})^{(n_i-1)/2} e^{-(n_i-1)s_i^2/2\sigma_i^2} \frac{1}{\sqrt{2\pi\delta_2^2}} e^{-(\ln(\sigma_i^2) - \underset{\sim}{x}_i'\underset{\sim}{\beta}_2)^2/2\delta_2^2}\right\}, \underset{\sim}{\theta} \in V.$$

Additional difficulties in the computation reside in this joint conditional posterior density. Observe that because $\underset{\sim}{\theta} \in V$, the $\theta_i$ are not independent, the $\sigma_i^2$ are not independent and $\theta_i$ and $\sigma_i^2$ are

not pairwise independent. However, note that the $\sigma_i^2$ are independent in their joint conditional posterior density, but the $\theta_i$ are not independent in their joint conditional posterior density. The $\sigma_i^2$ are drawn using the grid method with range $(\frac{1}{10}S_i^2, 10S_i^2)$, fairly wide, and the $\theta_i$ are drawn using Devroye's method.

For the Gibbs sampler, we used $2,500$ iterates as a burn-in and took every third iterate to get a random sample of $1,000$ iterates. We found that the Geweke tests for all the $\theta_i$ and the $\sigma_i^2$ are not significant and the effective sample sizes are all near the actual sample size of $1,000$ (mostly all of them are $1,000$). Therefore, we have an efficient Gibbs sampler and amazingly the computation took less than 20 seconds.

Next, we describe a slightly different computational method from the one described above. However, we just need to say how to draw samples from the conditional posterior densities of $(\underset{\sim}{\beta}_1, \delta_1^2)$ and $(\underset{\sim}{\beta}_2, \delta_2^2)$.

The conditional posterior density of $(\underset{\sim}{\beta}_2, \delta_2^2)$ is straight forward (i.e., we simply need to replace $S_i^2$ by $\sigma_i^2$). So that, letting $z_i = \ln(\sigma_i^2)$,

$$\underset{\sim}{\beta}_2 \mid \underset{\sim}{z}, \delta_2^2 \sim \text{Normal}\{\hat{\beta}_2, (X'X)^{-1}\delta_2^2\}, \delta_2^2 \mid \underset{\sim}{z} \sim \text{IG}\left\{\frac{n-p}{2}, \frac{\sum_{i=1}^n (z_i - \underset{\sim}{x}_i'\hat{\beta}_2)^2}{2}\right\}.$$

It is more difficult to sample the conditional posterior density of $(\underset{\sim}{\beta}_1, \delta_1^2)$,

$$\pi(\underset{\sim}{\beta}_1, \delta_1^2 \mid \underset{\sim}{\theta}, \underset{\sim}{\sigma^2}\underset{\sim}{\beta}_2, \delta_2^2, \mid D) \propto \frac{1}{\delta_1^2} \frac{\prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}_1)/\delta_1\}}{\int_{\underset{\sim}{\theta} \in V} \prod_{i=1}^{\ell} \phi\{(\theta_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}_1)/\delta_1\} d\underset{\sim}{\theta}}.$$

We started by using the Metropolis sampler. After we have used two different proposal densities, we found long-range dependence with low jumping rates, so we abandoned the Metropolis sampler. We decided to use grid samplers as follows. We fit the simpler model, where letting $X$ denote the $n \times p$ design matrix,

$$\underset{\sim}{\beta}_1 \mid \underset{\sim}{\hat{\theta}}, \delta_1^2 \sim \text{Normal}\{\hat{\beta}_1, (X'X)^{-1}\delta_1^2\}, \quad \delta_1^2 \mid \underset{\sim}{\hat{\theta}} \sim \text{IG}\left\{\frac{n-p}{2}, \frac{\sum_{i=1}^n (\hat{\theta}_i - \underset{\sim}{x}_i'\hat{\beta}_1)^2}{2}\right\},$$

with $\hat{\beta}_1 = (X'X)^{-1}X'\underset{\sim}{\hat{\theta}}$. Therefore, we can now sample $\underset{\sim}{\beta}_1$ and $\delta_1^2$ using the multiplication rule. Then, we find the posterior means (PM) and standard deviations (PSD) of each component of

$\underset{\sim}{\beta}_1$ and $\delta_1^2$; we choose their supports to be $PM \pm 6 * PSD$ with the lower bound for $\delta_1^2$ being $\max(0., PM - 6 * PSD)$. [Almost the entire support of a unimodal density is within this range; actually we have found the procedure to be nonsensitive to the choice of 6 to inference about the $\theta_1$]. We now run the grid method within the Gibbs sampler to draw $\underset{\sim}{\beta}_1$ and $\delta_1^2$ with the supports mentioned above for $\underset{\sim}{\beta}$ and $\delta_1^2$.

For the Gibbs sampler, we used $3,500$ iterates as a burn-in and took every fourth iterate to get a random sample of $1,000$ iterates. We found that the Geweke tests for all the $\theta_i$ and the $\sigma_i^2$ are not significant and the effective sample sizes are all near the actual sample size of $1,000$ (mostly all of them are $1,000$). Therefore, we have an efficient Gibbs sampler and amazingly the computation took less than 40 seconds. This is double the time (still fast) for the approximate Gibbs sampler above.

# APPENDIX C Discussions on Generalization

We show that the problem is more ubiquitous than we have stated in this paper. Then, we discuss issues with standard solutions using the logarithmic transformation. Recall that our problem is to provide estimates subjected to the lower bound inequality constraints and an equality benchmarking constraint. We discuss mainly the inequality constraint.

The Fay-Herriot model is

$$\hat{\theta}_i \mid \theta_i \overset{ind}{\sim} \text{Normal}(\theta_i, \hat{\sigma}_i^2),$$

$$\theta_i \mid \underset{\sim}{\beta}, \delta^2 \overset{ind}{\sim} \text{Normal}(\underset{\sim}{x}_i'\underset{\sim}{\beta}, \delta^2), i = 1, \ldots, \ell,$$

with prior $\pi(\underset{\sim}{\beta}, \delta^2)$. This is subjected to the inequality constraint, $\theta_i \geq c_i, i = 1, \ldots, \ell$, and the benchmarking constraint, $\sum_{i=1}^{\ell} \theta_i = a$, where $a$ is the target. Letting $\hat{\phi}_i = \hat{\theta}_i - c_i, i = 1, \ldots, \ell$, and $c = \sum_{i=1}^{\ell} c_i$. Then,

$$\hat{\phi}_i \mid \phi_i \overset{ind}{\sim} \text{Normal}(\phi_i, \hat{\sigma}_i^2), \tag{C.1}$$

$$\phi_i \mid \underset{\sim}{\beta}, \delta^2 \overset{ind}{\sim} \text{Normal}(\underset{\sim}{x}_i'\underset{\sim}{\beta}, \delta^2), \phi_i \geq 0, i = 1, \ldots, \ell, \tag{C.2}$$

with $\sum_{i=1}^{\ell} \phi_i = a - c$; note that there is a change in the regression coefficients. Therefore, we have a general problem with positivity constraints and a benchmarking constraint, and the problem is not specific to agriculture. The solution of problem remains the same as we have done in this paper, but we can use the logarithmic transformation to avoid the positivity constraint.

There are two ways to proceed without the positivity constraints.

a. Transform the $\hat{\phi}_i$, replacing $\hat{\phi}_i$ by $\log(\hat{\phi}_i)$ in (C.1). Note that some of the $\hat{\phi}_i$ can be negative, thereby loosing some generality. For the case when they are positive, we can approximate the means and the variances of the normal distribution in (C.1) using a first-order Taylor's series approximation. That is, $\log(\hat{\phi}_i) \mid \phi_i \overset{ind}{\sim} \text{Normal}(\log(\phi_i), \frac{\hat{\sigma}_i^2}{\hat{\phi}_i^2})$. One can proceed in (C.2) with either a log-normal regression or another distribution for positive size data (e.g., gamma regression).

b. Transform the $\phi_i$, replacing $\phi_i$ by $e^{\phi_i}$ in (C.1). This introduces non-conjugacy with (C.2), thereby creating difficulties in computation.

Note again that benchmarking is done in an output analysis as we have done in this paper, and both single shrinkage models and double shrinkage models can be done. When the logarithmic transformation is used, back transformation to the original $\phi_i$ is problematic (e.g., Manandhar and Nandram, 2021). However, the methodology in this paper provides our front line solution.

## ACKNOWLEDGEMENT AND DISCLAIMERS

Dr. Cruze's and Dr. Erciulescu's contributions to this research were made during their tenure in the Research and Development Division at USDA's NASS. The findings and conclusions in this paper are those of the authors and should not be construed to represent any official USDA or US Government determination or policy. This research was supported by the U.S. Department of Agriculture, National Agriculture Statistics Service. Balgobin Nandram was supported by a grant from the Simons Foundation (#353953, Balgobin Nandram).

## References


Battese, G. E., Harter, R. 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 (401), 28-36.

Chen, L., Cruze, N. B. and Nandram, B. (2022), "Hierarchical Bayesian Model with Inequality Constraints for US County Estimates," *Journal of Official Statistics* (to appear).

Chen, X. and Nandram, B. (2022), "Bayesian Inference for Multinomial Data from Small Areas Incorporating Uncertainty about Order Restriction," *Survey Methodology*, 48 (1), 145-176.

Cruze, N. B., Erciulescu, A. L., Nandram, B., Barboza, W. J. and Young, L. J. (2019), "Producing Official County-Level Agricultural Estimates in the United States: Needs and Challenges," *Statistical Science*, 34 (2), 301-316.

Dass, S. C., Maiti, T., Ren, H. and Sinha, S. (2012), "Confidence Interval Estimation of Small Area Parameters Shrinking both Means and Variances," *Survey Methodology*, 38 (2), 173-187.

Devroye, L. (1986), *Non-Uniform Random Variate Generation*, Springer-Verlag: New York.

Erciulescu, A. L., Cruze, N. B. and Nandram, B. (2018), "Benchmarking a triplet of official estimates," *Environmental and Ecological Statistics*, 25, 523-547.

Erciulescu, A. L. Cruze, N. B. and Nandram, B. (2019), "Model-based county level crop estimates incorporating auxiliary sources of information," *Journal of the Royal Statistical Society*, Series A, 182 (1), 283-303.

Erciulescu, A. L. Cruze, N. B. and Nandram, B. (2020), "Statistical Challenges in Combining Survey and Auxiliary Data to Produce Official Statistics," *Journal of Official Statistics*, 36 (1), 63-88.

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.

Genz, A. (1992), "Numerical Computation of Multivariate Normal Probabilities," *Journal of Computational and Graphical Statistics*, 1, 141-149.

Geweke, J. (1991), "Efficient Simulation from the Multivariate normal and Student-t Distributions Subject to Linear Constraints and the Evaluation of Constraint Probabilities," *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*, 23, 571-578.

Gonzalez-Manteiga, W., Lombardia, M. J., Molina, I., Morales, D. and SantaMaria, L. (2010), "Small Area Estimation under Fay-Herriot Models with Nonparametric Estimation of Heteroscedasticity," *Statistical Modelling*, 10, 215–239.

Janicki, R. and Vesper, A. (2017), "Benchmarking Techniques for Reconciling Bayesian Small Area Models at Distinct Geographical Levels," *Statistical Methods and Applications*, 26, 557-581.

Maiti, T., Ren, H. and Sinha, S. (2014), "Prediction Error of Small Area Predictors Shrinking Both Means and Variances," *Scandinavian Journal of Statistics*, 41, 775-790.

Manandhar, B. and Nandram, B. (2021), "Hierarchical Bayesian models for continuous and positively skewed data from small areas," *Communications in Statistics - Theory and Methods*, 50 (4), 944-962.

Molina, I., Nandram, B. and Rao, J. N. K. (2014), "Small Area Estimation of General Parameters with Application to Poverty Indicators: A Hierarchical Bayes Approach," *The Annals of Applied Statistics*, 8 (2), 852-885.

Nandram, B. and Erhardt, E. B. (2004), "Fitting Bayesian Two-Stage Generalized Linear Models Using Random Samples via the SIR Algorithm," *Sankhya: The Indian Journal of Statistics*, 66 (4), 733-755.

Nandram, B. and Sayit, H. (2011), "A Bayesian analysis of small area probabilities under a constraint," *Surev Methodology*, 37, 137-152.

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

Nandram, B., Erciulescu and Cruze, N. (2019), "Bayesian Benchmarking of the Fay-Herriot Model Using Random Deletion," *Survey Methodology*, 45 (2), 365-390.

Nandram, B., Sedransk, J. and Smith, S. J. (1997), "Order Restricted Bayesian Estimation of the Age Composition of a Population of Atlantic Cod," *Journal of the American Statistical Association*, 92, 33-40.

National Academies of Sciences, Engineering, and Medicine (2017), *Improving Crop Estimates by Integrating Multiple Data Sources*, The National Academies Press, Washington, DC. https://doi.org/10.17226/24892.

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

Ridgway, J. (2016), "Computation of Gaussian Orthant Probabilities in High Dimension," *Statistics and Computing*, 26, 899-916.

Sen, P. K. and Silvapulle, M. J. (2002), "An Appraisal of Some Aspects of Statistical Inference Under Inequality Constraints," *Journal of Statistical Planning and Inference*, 107, 3-43.

Silvapulle, M. J. and Sen, P. K. (2005), *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*, Wiley: New York.

Steorts, R. C., Schmid, T. and Tzavidis (2020), "Smoothing and Benchmarking for Small Area Estimation," *International Statistical Review*, 88 (3), 580-598.

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. and Fuller, W. A. (2003), "The Mean Squared Error of Small Area Predictors Constructed with Estimated Area Variances," *Journal of the American Statistical Association*, 98, 716–723.

You, Y. and Chapman, B. (2006), "Small Area Estimation using Area Level Models and Estimated Sampling Variances," *Survey Methodology*, 32, 97–103.