Hierarchical Bayesian Model for County-Level Cash Rental Rates

Lu Chen* Balgobin Nandram†

Abstract

Small area models have gained increased attention by statistical agencies. They can "borrow strength" from related areas across space and/or time or through auxiliary information and they can provide "indirect" but reliable estimates for small areas with small or even zero sample sizes while also increasing the precision. The United States Department of Agriculture's (USDA's) National Agricultural Statistics Service (NASS) conducts the Cash Rents Survey (CRS) to provide the basis for county estimates of the cash rent paid for irrigated cropland, non-irrigated cropland, and pasture land. Estimates of cash rental rates are useful to farmers, economists, and policy makers. However, realized sample sizes at the county level are often too small to support reliable direct estimates. To improve the less reliable direct estimates, model-based estimates have been extensively discussed in the literature. We propose a hierarchical Bayesian (HB) area-level two-component mixture model to account for outliers that incorporates two years of data with a discounting factor for the first year. When compared to the standard HB method based on normality assumptions, the proposed method to handle outliers is robust. In addition, it is a general model that puts the two years of data together and it avoids correlations by using a power prior that partly discounts past data. A 2016 and 2017 case study illustrates the improvement of the direct survey estimates for areas with small sample sizes by using auxiliary information and by borrowing information across areas.

Key Words: Block Gibbs sampler, Grid method, Mixture model, Power prior, Outliers, Small area estimation, Survey data

*National Institute of Statistical Sciences and USDA National Agricultural Statistics Service, 1400 Independence Avenue, SW, Room 6412 B, Washington, DC 20250-2054. Email: lchen@niss.org.

†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

1. Introduction

The United States Department of Agriculture's (USDA's) National Agricultural Statistics Service (NASS), one of thirteen US Federal Statistical Agencies, provides official statistics of the county-level cash rental rates for irrigated cropland, non-irrigated cropland, and pasture land. Estimates of cash rental rates are useful to farmers, economists, and policy makers. However, realized sample sizes at the county level are often too small to support reliable direct estimates. Traditionally, the NASS Agricultural Statistics Board has relied on expert opinion to produce official statistics using the survey estimates as a foundation informed by other auxiliary information, including historical data, reliable administrative data, and other non-survey data. Although external reviews have consistently found that NASS estimates are the gold standard for the agricultural industry, the process lacked transparency and reproducibility and did not lead to valid measures of uncertainty. In 2014, NASS entered into a cooperative agreement with the Committee on National Statistics (CNSTAT) to review the NASS county estimates programs, crops county estimates and cash rents county estimates. In its consensus report, the CNSTAT panel recommended that NASS transition to model-based estimates.

Small area models have gained increased attention by statistical agencies and offices around the world. Model-based approaches in small area estimation have been shown to be useful in producing reliable small area estimates. Small area models can "borrow strength" from related areas or through auxiliary information to provide "indirect" but reliable estimates for areas with small or even zero sample sizes while also increasing the precision. Two major types of small area models, area-level and unit-level models, have been developed based on both frequentist and Bayesian methods. Rao and Molina (2015) provides a comprehensive overview of the development, methods and applications of small area estimation, including various types of area-level and unit-level models. The most common model in small area estimation is the Fay-Herriot (FH) model by Fay and Herriot (1979). It is an area-level model based on the normality assumption; that is, the direct estimates and area-level random effects are each assumed to follow a normal distribution. Battese et al. (1988) proposed the unit-level model, nested-error regression (NER) model, when data are available on the individual sampled units. It is developed under the assumption of normality of the small area random effects and unit-level random errors.

USDA's NASS has explored different model-based approaches for county-level cash rental rate estimates. In a frequentist framework, Berg et al. (2014) propose a univariate area-level model that involves

fitting two sets of survey-based direct estimates. Berg et al. (2014) assume that the variances for the two years are the same. The model can be characterized as an extension of the FH model. Erciulescu et al. (2019) propose a HB bivariate unit-level model under the normality assumptions. The model is flexible to allow the variances to differ between the two time-points. However, it is computationally intensive to fit the models in production. Both models use information from two survey years to gain in efficiency due to the fact that the rental rates from the current and previous year are often highly correlated despite differences between the sets of respondents in both years.

Because the interest for NASS programs is in constructing summaries for different levels of geography (county, state, regional, and U.S. levels), the Bayesian approach to model fitting and estimation is preferable. Recent studies and papers related to NASS small area estimation research on county estimates of crops and farm labor program have shown that the HB small area models can incorporate auxiliary sources of data with survey estimates to improve the precision and increase the accuracy of related NASS official estimates. Nandram et al. (2022) and Chen et al. (2022b) proposed and implemented HB subarea-level models with inequality constraints to produce county-level estimates that satisfy important relationships between the estimates and administrative data, along with the associated measures of uncertainty. Chen et al. (2022a) discussed several HB subarea-level models in support of estimates of interest in the Farm Labor Survey. The resulting framework provided a complete set of coherent estimates for all required geographic levels and the modeling process was incorporated into the official Farm Labor publication for the first time in 2020.

Our study addresses the issues of constant variances assumption in two years and the issues of computation by proposing a general HB model that puts the two years of data together and avoids the two-year correlations by using a power prior (Chen and Ibrahim (2000)) that partly discounts past data. In addition, outliers issues are taken into account to increase robustness for the standard HB area-level model with the normality assumptions. One way to accommodate outliers is to use a two-component mixture model (e.g., Gershunskaya and Lahiri (2017)). They use an Empirical Bayes approach assisted by the expectation-maximization (EM) algorithm. Chakraborty et al. (2019) and Goyal et al. (2021) have provided a full Bayesian approach for the unit-level nested error regression model. By using a two-component mixture of normal distributions, this model accommodates populations where a small portion of unit-level errors come from a secondary distribution with a larger variance than the primary distribution. In practice, because unit-level models generally require substantially more computational time, area-level models are more applicable for the production of official statistics that are published on tight timelines. Therefore, we focus on

the area-level two-component mixture model in this paper.

The paper is structured as follows. Section 2 describes the survey procedures and background. The proposed HB two-component mixture model with technical details is presented in Section 3. In Section 4, a case study using 2016 and 2017 cash rent data illustrates the performance of the model. Section 5 provides a summary and some discussion of possible future research for cash rental rates and small area estimation more generally.

2. Data Sources and Requirements

2.1 Survey Data

The Cash Rents Survey is conducted on an annual basis. The survey obtains acres rented and cash rental rates from a statistically representative sample of farmers and ranchers in the United States, excluding Alaska. This survey provides the basis for estimates of the current year's cash rents paid for irrigated cropland, non-irrigated cropland, and permanent pastureland. From 1950 to 1974, a list survey of real estate appraisers was used to estimate state-level cash rents. Beginning in 1974, producers provided information about their rental agreements by responding to questions on the June Area Survey. In the 2008 farm bill, NASS was mandated to provide mean rental rates for all counties (not just states) with at least 20,000 acres of crop land.

The target population for the cash rents estimate program is all farms and ranches with $1,000 or more in agricultural sales (or potential sales) who rent land from others on a cash rent basis. The Cash Rent Survey sample is selected from a list frame of farm and ranch operators maintained by NASS. NASS is constantly seeking qualifying farming operations from outside sources to be added to the list. A profile, known as control data, of each operation is maintained which indicates what the farm has historically produced and a general indication of size. This information allows NASS to define sampling populations that are specific to each survey and employ advanced and more efficient sample designs. Samples for the Cash Rents Survey are drawn with a county level stratified design to produce state and county level estimates. Large operations in each county are stratified into the census strata, where all are included in the sample. The national sample size for the Cash Rents Survey is approximately 260,000. The sample is stratified by state and county within state to produce state and county-level estimates. Data collection occurs from late February until the end of June. Variances for cash rental rate estimates are constructed using a second-order Taylor series expansion

for the ratio.

From the Cash Rents Survey, county, state, and national rental rates (dollars/acre) for each land-use category (irrigated, non-irrigated, and pasture) are published (see Figure 1 for the 2020 county-level published cash rental rate estimates for pasture land). Although total value of cash rents and acres rented on a cash basis are computed, these values have not been published. Historically, the direct survey estimates were reviewed and, if deemed appropriate, adjusted by NASS staff or the Agricultural Statistics Board. The primary reason for adjustment was a large difference between the previous year's published cash rental rate and the current year's survey estimate. The adjusted estimate was restricted to being between, or on, the current survey estimate and the previous year's published estimate so that the direction of change was honored. If the number of responses within a county was substantial, then the survey estimate received the greatest weight. If only a small number of responses were received, then the previous year's published value and the estimates from surrounding counties or the agricultural statistics district were given more weight. Any model to replace the expert opinion needs to follow the guidelines used in the review process.

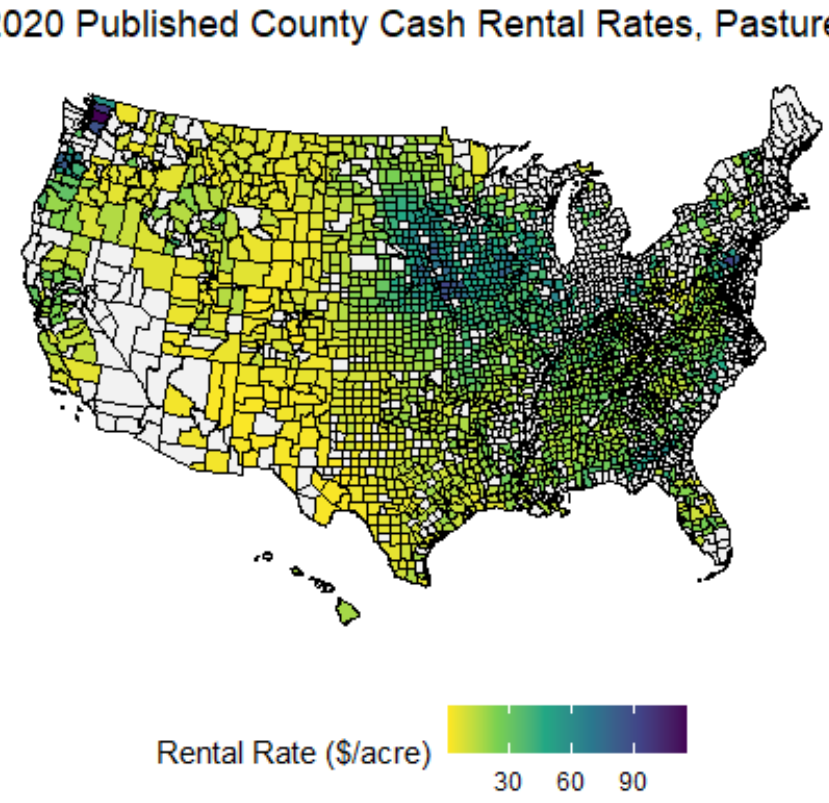

Figure 1: County-level published cash rental rate estimates for pasture land

3. Model

In this section, we describe a two-component mixture model that accounts directly for outliers and it uses the power prior to input how much of the past data one wants to use. One way to accommodate outlier operations is to use a two-component mixture model. In addition, if research interest is only in the second year (i.e., the year of the survey), we can use the data in the prior year(s) to obtain a power prior (Ibrahim and Chen, 2000; Ibrahim, Chen, Gwon and Chen, 2015) for the parameters of the current year. In this way, the correlation between survey data from the two consecutive years can be avoided, which reduces the computational time to a manageable scale.

3.1 A Hierarchical Bayesian Two-Component Mixture Model

For modeling, the subscript 1 will denote year 1 (e.g., previous year), and the subscript 2 will denote year 2 (e.g., the current year). Let the counties $1, \ldots, \ell_1, \ldots, \ell_1 + \ell_c$ on the first year and $\ell_1 + 1, \ldots, \ell_1 + \ell_c + \ell_2$ on the second year. That is, there are ℓ_1 counties sampled only on the first year, ℓ_c sampled on both years and ℓ_2 sampled only on the second year; ℓ_1 and ℓ_2 are usually very small relative to ℓ_c. Let $\underset{\sim}{x}_{tij}$ denote the vector of covariates; let y_{tij} denote the direct estimates and $\hat{\sigma}^2_{ti}$ denote the corresponding sampling variances, where $i = 1, \ldots, \ell_1 + \ell_c$ for $t = 1$ and $i = \ell_1 + 1, \ldots, \ell_1 + \ell_c, \ldots, \ell_1 + \ell_c + \ell_2$ for $t = 2$.

For the first year, letting $c = \ell_1 + \ell_c$,

$$\hat{\theta}_{1i} \mid \theta_{1i}, a, p, \rho \overset{ind}{\sim} (1-p)\text{Normal}\{\theta_{1i}, \frac{\rho\hat{\sigma}^2_{1i}}{a}\} + p\text{Normal}\{\theta_{1i}, \frac{\hat{\sigma}^2_{1i}}{a}\}, i = 1, \ldots, c,$$

and for the second year, letting $\ell = \ell_1 + \ell_c + \ell_2$,

$$\hat{\theta}_{2i} \mid \theta_{2i}, a, p, \rho \overset{ind}{\sim} (1-p)\text{Normal}\{\theta_{2i}, \rho\hat{\sigma}^2_{2i}\} + p\text{Normal}\{\theta_{2i}, \hat{\sigma}^2_{2i}\}, i = \ell_1 + 1, \ldots, \ell,$$

where $0 < a, 2p, \rho < 1$. Here, the number of outliers is assumed to be fewer than the number of non-outliers. Note that there are a set of common counties with subscripts, $\ell_1 + 1, \ldots, c$. Then, the positivity constraints, with the covariates, are included as

$$\theta_{1i} \mid \underset{\sim}{\beta}, \delta^2 \overset{ind}{\sim} \text{Normal}(\underset{\sim}{x}'_{1i}\underset{\sim}{\beta}, \delta^2), \theta_{1i} > 0, i = 1, \ldots, c,$$

and

$$\theta_{2i} \mid \underset{\sim}{\beta}, \delta^2 \overset{ind}{\sim} \text{Normal}(\underset{\sim}{x}'_{2i}\underset{\sim}{\beta}, \delta^2), \theta_{2i} > 0, i = \ell_1 + 1, \ldots, \ell.$$

Finally, a priori, we assume

$$\pi(\underset{\sim}{\beta}, \delta^2) \propto \frac{1}{(1+\delta^2)^2}, \underset{\sim}{\beta} \in R^p, \delta^2 > 0.$$

The prior distributions for the θ_{1i} and the θ_{2i} can be written as follows,

$$\pi(\theta_{1i} \mid \underset{\sim}{\beta}, \delta^2) = \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{1}{2\delta^2}(\theta_{1i} - \underset{\sim}{x}'_{1i}\underset{\sim}{\beta})^2} / \Phi(\frac{\underset{\sim}{x}'_{1i}\underset{\sim}{\beta}}{\sqrt{\delta^2}}), \theta_{1i} > 0, i = 1, \ldots, c,$$

and

$$\pi(\theta_{2i} \mid \underset{\sim}{\beta}, \delta^2) = \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{1}{2\delta^2}(\theta_{2i} - \underset{\sim}{x}'_{2i}\underset{\sim}{\beta})^2} / \Phi(\frac{\underset{\sim}{x}'_{2i}\underset{\sim}{\beta}}{\sqrt{\delta^2}}), \theta_{2i} > 0, i = \ell_1 + 1, \ldots, \ell.$$

For (p, a, ρ), we assume

$$\pi(p, a, \rho) \propto 1, 0 < a, 2p, \rho < 1$$

with independent priors on these components.

Let D denote the set of data values, $D = \{\hat{\underset{\sim}{\theta}}_1, \hat{\underset{\sim}{\theta}}_2, \hat{\underset{\sim}{\sigma}}_1^2, \hat{\underset{\sim}{\sigma}}_2^2\}$. Using Bayes' theorem, the joint posterior density is

$$\begin{aligned}
\pi(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2, \underset{\sim}{\beta}, \delta^2, p, a, \rho \mid D) \propto\ & \pi(\underset{\sim}{\beta}, \delta^2)\pi(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2 | \underset{\sim}{\beta}, \delta^2) \\
& \times \prod_{i=1}^{c} \left\{ (1-p)\text{Normal}(\underset{\sim}{x}'_{1i}\underset{\sim}{\beta} + \nu_i, \rho\frac{\hat{\sigma}_{1i}^2}{a}) + p\text{Normal}(\underset{\sim}{x}'_{1i}\underset{\sim}{\beta} + \nu_i, \frac{\hat{\sigma}_{1i}^2}{a}) \right\} \\
& \times \prod_{i=\ell_1+1}^{\ell} \left\{ (1-p)\text{Normal}(\underset{\sim}{x}'_{2i}\underset{\sim}{\beta} + \nu_i, \rho\hat{\sigma}_{2i}^2) + p\text{Normal}(\underset{\sim}{x}'_{2i}\underset{\sim}{\beta} + \nu_i, \hat{\sigma}_{2i}^2) \right\}.
\end{aligned}$$

It is computationally advantageous to augment the joint posterior density using the latent binary variables, z_{ti}, where $z_{ti} = 1$ if a county is an outlier and $z_{ti} = 0$ otherwise. Then,

$$\begin{aligned}
\pi(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2, \underset{\sim}{\beta}, \delta^2, p, a, \rho, \underset{\sim}{z} \mid D) \propto\ & \pi(\underset{\sim}{\beta}, \delta^2)\pi(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2 | \underset{\sim}{\beta}, \delta^2) \\
& \times \prod_{i=1}^{c} \left\{ [(1-p)\text{Normal}(\underset{\sim}{x}'_{1i}\underset{\sim}{\beta} + \nu_i, \rho\frac{\hat{\sigma}_{1i}^2}{a})]^{1-z_{1i}} [p\text{Normal}(\underset{\sim}{x}'_{1i}\underset{\sim}{\beta} + \nu_i, \frac{\hat{\sigma}_{1i}^2}{a})]^{z_{1i}} \right\} \\
& \times \prod_{i=\ell_1+1}^{\ell} \left\{ [(1-p)\text{Normal}(\underset{\sim}{x}'_{2i}\underset{\sim}{\beta} + \nu_i, \rho\hat{\sigma}_{2i}^2)]^{1-z_{2i}} [p\text{Normal}(\underset{\sim}{x}'_{2i}\underset{\sim}{\beta} + \nu_i, \hat{\sigma}_{2i}^2)]^{z_{2i}} \right\}
\end{aligned}$$

Denote the set of parameters as $\Omega = (\Omega_1, \Omega_2, \underset{\sim}{\theta})$, with $\Omega_1 = \{a, p, \rho, \underset{\sim}{z}\}$, $\Omega_2 = \{\underset{\sim}{\beta}, \delta^2\}$, and $\underset{\sim}{\theta} = \{\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2\}$. For $\theta_{1i} > 0, i = 1, \ldots, c, \theta_{2i} > 0, i = \ell_1 + 1, \ldots, \ell$, the joint posterior density is

$$\begin{aligned}\pi(\Omega \mid D) \propto & \frac{1}{(1+\delta^2)^2} \prod_{i=1}^{c} \left\{ (1-p)\sqrt{\frac{a}{2\pi\rho\hat{\sigma}_{1i}^2}} e^{-\frac{a}{2\rho\hat{\sigma}_{1i}^2}(\theta_{1i}-\hat{\theta}_{1i})^2} \right\}^{1-z_{1i}} \left\{ p\sqrt{\frac{a}{2\pi\hat{\sigma}_{1i}^2}} e^{-\frac{a}{2\hat{\sigma}_{1i}^2}(\theta_{1i}-\hat{\theta}_{1i})^2} \right\}^{z_{1i}} \\ & \times \prod_{i=\ell_1+1}^{\ell} \left\{ (1-p)\sqrt{\frac{1}{2\pi\rho\hat{\sigma}_{2i}^2}} e^{-\frac{1}{2\rho\hat{\sigma}_{2i}^2}(\theta_{2i}-\hat{\theta}_{2i})^2} \right\}^{1-z_{2i}} \left\{ p\sqrt{\frac{1}{2\pi\hat{\sigma}_{2i}^2}} e^{-\frac{1}{2\hat{\sigma}_{2i}^2}(\theta_{2i}-\hat{\theta}_{2i})^2} \right\}^{z_{2i}} \\ & \times \prod_{i=1}^{c} \left\{ \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{1}{2\delta^2}(\theta_{1i}-\underset{\sim}{x}_{1i}'\underset{\sim}{\beta})^2} / \Phi\left(\frac{\underset{\sim}{x}_{1i}'\underset{\sim}{\beta}}{\sqrt{\delta^2}}\right) \right\} \prod_{i=\ell_1+1}^{\ell} \left\{ \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{1}{2\delta^2}(\theta_{2i}-\underset{\sim}{x}_{2i}'\underset{\sim}{\beta})^2} / \Phi\left(\frac{\underset{\sim}{x}_{2i}'\underset{\sim}{\beta}}{\sqrt{\delta^2}}\right) \right\}. \end{aligned} \tag{1}$$

The joint posterior density is intractable and thus the standard Gibbs sampler cannot be applied. The computational method we proposed is discussed in the next section. Note that not-in-sample prediction is not needed in the small area model for our particular application. The primary interest is the posterior density of the θ_{2i}. In addition, for benchmarking, we need to obtain posterior inference about $\theta_{2i}, i = \ell_1 + 1, \ldots, \ell$ subject to the benchmarking constraint with a specified target. For convenience we re-label the θ_{2i} as $\theta_i, i = 1, \ldots, n$, where n is the number of counties available on the second year. For the i^{th} county, let a_i denote the total number of acres under cash rents. Then, we need $\sum_{i=1}^{n} \frac{a_i}{\sum_{i=1}^{n} a_i}\theta_i = T$, where T is the pre-published estimate of cash rents at the state level. Benchmarking is usually obtained from the output analysis using raking, relatively simpler than other methods.

3.2 Computation

The posterior density (1) is a non-standard multivariate density, and there are difficulties in fitting it using MCMC methods. We can rewrite the joint posterior density as

$$\pi(\Omega \mid D)) = \pi(\Omega_1, \Omega_2, \underset{\sim}{\theta} \mid D)) \propto \pi(\Omega_1, \Omega_2 | D)\pi(\underset{\sim}{\theta} | \underset{\sim}{\beta}, \delta^2, \Omega_1, D)$$

We first integrate out the $\underset{\sim}{\theta}$ from the joint posterior density. Since there are many of parameters in this model, and each of them is best done using a grid method, we can save a lot of time not drawing them inside the Gibbs sampler. We use a two-block Gibbs sampler to draw samples from the posterior density of $\pi(\underset{\sim}{\theta} | \underset{\sim}{\beta}, \delta^2, \Omega_1, D)$.

First, we define the following shrinkage parameters. For the first year, corresponding to the first and

second components of the mixture model respectively, we define

$$\omega_{1i} = \frac{\delta^2}{\delta^2 + \rho\hat{\sigma}_{1i}^2/a}, \;\; \omega_{1i}^* = \frac{\delta^2}{\delta^2 + \hat{\sigma}_{1i}^2/a}, i = 1, \ldots, c,$$

and for the current year, we define

$$\omega_{2i} = \frac{\delta^2}{\delta^2 + \rho\hat{\sigma}_{2i}^2}, \;\; \omega_{2i}^* = \frac{\delta^2}{\delta^2 + \hat{\sigma}_{2i}^2}, i = \ell_1 + 1, \ldots, \ell.$$

The introduction of the latent variables creates a simplified and more easily sampled (not completely) joint posterior density. In fact, the joint posterior density is

$$\begin{aligned}
\pi(\Omega \mid D) \propto & \frac{1}{(1+\delta^2)^2} \\
& \times \prod_{i=1}^{c} \left\{ (1-p)\sqrt{\frac{\omega_{1i}}{2\pi\delta^2}} e^{-\frac{\omega_{1i}}{2\delta^2}(\hat{\theta}_{1i} - \underset{\sim}{x}_{1i}'\underset{\sim}{\beta})^2} \frac{\Phi(\frac{\omega_{1i}\hat{\theta}_{1i} + (1-\omega_{1i})\underset{\sim}{x}_{1i}'\underset{\sim}{\beta}}{\sqrt{(1-\omega_{1i})\delta^2}})}{\Phi(\frac{\underset{\sim}{x}_{1i}'\underset{\sim}{\beta}}{\sqrt{\delta^2}})} \right\}^{1-z_{1i}} \\
& \times \prod_{i=1}^{c} \left\{ p\sqrt{\frac{\omega_{1i}^*}{2\pi\delta^2}} e^{-\frac{\omega_{1i}^*}{2\delta^2}(\hat{\theta}_{1i} - \underset{\sim}{x}_{1i}'\underset{\sim}{\beta})^2} \frac{\Phi(\frac{\omega_{1i}^*\hat{\theta}_{1i} + (1-\omega_{1i}^*)\underset{\sim}{x}_{1i}'\underset{\sim}{\beta}}{\sqrt{(1-\omega_{1i}^*)\delta^2}})}{\Phi(\frac{\underset{\sim}{x}_{1i}'\underset{\sim}{\beta}}{\sqrt{\delta^2}})} \right\}^{z_{1i}} \\
& \times \prod_{i=\ell_1+1}^{\ell} \left\{ (1-p)\sqrt{\frac{\omega_{2i}}{2\pi\delta^2}} e^{-\frac{\omega_{2i}}{2\delta^2}(\hat{\theta}_{2i} - \underset{\sim}{x}_{2i}'\underset{\sim}{\beta})^2} \frac{\Phi(\frac{\omega_{2i}\hat{\theta}_{2i} + (1-\omega_{2i})\underset{\sim}{x}_{2i}'\underset{\sim}{\beta}}{\sqrt{(1-\omega_{2i})\delta^2}})}{\Phi(\frac{\underset{\sim}{x}_{2i}'\underset{\sim}{\beta}}{\sqrt{\delta^2}})} \right\}^{1-z_{2i}} \\
& \times \prod_{i=\ell_1+1}^{\ell} \left\{ p\sqrt{\frac{\omega_{2i}^*}{2\pi\delta^2}} e^{-\frac{\omega_{2i}^*}{2\delta^2}(\hat{\theta}_{2i} - \underset{\sim}{x}_{2i}'\underset{\sim}{\beta})^2} \frac{\Phi(\frac{\omega_{2i}^*\hat{\theta}_{2i} + (1-\omega_{2i}^*)\underset{\sim}{x}_{2i}'\underset{\sim}{\beta}}{\sqrt{(1-\omega_{2i}^*)\delta^2}})}{\Phi(\frac{\underset{\sim}{x}_{2i}'\underset{\sim}{\beta}}{\sqrt{\delta^2}})} \right\}^{z_{2i}} ;
\end{aligned}$$

see Appendix A for details.

We use the two-block Gibbs sampler to sample the joint posterior density. Specifically, we sample $\Omega_1 \mid \Omega_2, D$ and $\Omega_2 \mid \Omega_1, D$, each in turn, until convergence. Note that this joint posterior density does not contain θ_{ti} and this is enormous gain because they are not included in the Gibbs sampler, thereby providing a more efficient Gibbs sampler. We will sample the θ_{ti} in an output analysis (i.e., after the Gibbs sampler is finished).

Let

$$d_{1i}^2 = \frac{a}{\rho\hat{\sigma}_{1i}^2}\{(1-z_{1i})+z_{1i}\rho\}, i=1,\ldots,c,\ d_{2i}^2 = \frac{1}{\rho\hat{\sigma}_{2i}^2}\{(1-z_{2i})+z_{2i}\rho\}, i=\ell_1+1,\ldots,\ell.$$

It is worth noting that d_{1i} and d_{2i} do not depend on Ω_2. Conditioning on $\underset{\sim}{\beta}, \delta^2$ and Ω_1, the θ_{1i} and the θ_{2i} are all independent. Letting $\lambda_{1i} = \frac{\delta^2 d_{1i}^2}{1+\delta^2 d_{1i}^2}, i=1,\ldots,c$, and $\lambda_{2i} = \frac{\delta^2 d_{2i}^2}{1+\delta^2 d_{2i}^2}, i=\ell_1+1,\ldots,\ell$, we have

$$\theta_{ti} \mid \underset{\sim}{\beta}, \delta^2, \Omega_1, D \overset{ind}{\sim} \text{Normal}\{\lambda_{ti}\hat{\theta}_{ti} + (1-\lambda_{ti})\underset{\sim}{x}_{ti}'\underset{\sim}{\beta}, (1-\lambda_{ti})\delta^2\}, \theta_{ti}>0, i=1,\ldots,c,\ t=1,2. \quad (2)$$

4. Case Study of 2016 and 2017 Cash Rental Rates

In this section, 2016 and 2017 cash rental rate data are selected as the case study. We fit the cash rental rate model for each land type (non-irrigated land, irrigated land and pasture land) for twelve regions in US. NASS has twelve regional offices across the country, each of which is responsible for the statistical work in several states that are close to each other.

$\underset{\sim}{x}_{ti}$ are the known auxiliary information used in the model and include an intercept, the corresponding previous year county-level official estimates, the number of positive responses, and The National Commodity Crop Productivity Indices (NCCPIs). NCCPIs, which measure the quality of the soil for growing non-irrigated crops in climate conditions best suited for various crops, are available at the county level in the US. They are also correlated to the crop yield.

In Section 4.1, we discuss the model fit, Bayesian diagnostics, and the computation time related to the model. In Section 4.2, we show nationwide comparisons among model, survey and the published estimates.

4.1 Model Fit and Estimation

The two-component mixture model introduced in Section 3 is a useful tool for producing model-based estimates of cash rental rates with measures of uncertainty. We fit the cash rental rate model for each land type (non-irrigated land, irrigated land and pasture land) for twelve regions in the US.

Convergence diagnostics are conducted. The convergence for parameters involved in the two-block Gibbs samplers (Ω_1, Ω_2) is monitored using trace plots, the Geweke test of stationarity(Geweke (1992)) and the effective sample sizes. There is a single run of 10,000 iterates to fit the model and a "burn in" of 2000 iterates. In order to eliminate the correlations among neighboring iterations, those iterations are

thinned by taking a systematic sample of 1 in every 8 samples. Finally, 1000 MCMC samples are obtained to construct the posterior distributions of parameters and make inferences for the current year cash rental rate estimates $\theta_{2i},\ i = \ell_1 + 1, \ldots, \ell$. We find that the Geweke tests for all the parameters in the model are not significant and the effective sample sizes are all near the actual sample size of 1000 (mostly all of them are 1000). Computation time is an additional factor when candidate models are evaluated for use in NASS production. The described model and their posterior summaries, which generate estimates and measures of uncertainty for thousands of counties per land type, must be completed in less than 1 day during the annual production windows. In the case study, the computation time for models fitted for twelve regions by three land types was less than 2 hours. This is acceptable in the NASS production process.

4.2 Numerical Summaries and Comparisons

The model of Section 3 was fit to the cash rental rates reported on the 2016 and 2017 Cash Rent Surveys (CRSs) at the county level for twelve regions in the US. We present in Figure 2 a graphical display of published estimates v.s. survey estimates and model estimates, respectively. The horizontal axes of the plots represent the 2017 nationwide county-level survey estimates (left panel) or the mixture model estimates (right panel) by three land types. The vertical axes represent the published estimates. It is straightforward to see that the points in the right panel plots of the published estimates versus the model estimates are scattered around the 45 degree line. In the left panel, some of the survey estimates are far away from the 45 degree line, especially points corresponding to the pasture land (V).

To evaluate the effectiveness of the estimator, we computed the following two deviation measures for the model estimates and survey estimates from the "truth" (i.e. the published estimates): the average absolute relative deviation (AARD) and the average squared relative deviation (ASRD). External evaluation of potential models can shed light on their usefulness. We conducted external model evaluation at the time of research using the previous published estimates. Note that the published estimates would not be available for the current year. However, they are appropriate for use in assessing the quality and reasonableness of model-based estimates at the research stage.

- Average absolute relative deviation (AARD):

$$AARD = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{\theta}_i - \theta_i^{pub}|}{\theta_i^{pub}};$$

- Average squared relative deviation (ASRD):

$$ASRD = \frac{1}{n}\sum_{i=1}^{n}\frac{(\hat{\theta}_i - \theta_i^{pub})^2}{\theta^{pub_i^2}},$$

where n is the number of all counties and $\hat{\theta}_i$ is the county-level after-benchmarking posterior means and θ_i^{pub} is the published estimates. The number of responses for cash rental rates varies with county in each state. Small area models tend to improve the accuracy of estimates comparing to the accuracy of survey estimates, especially in areas with small sample sizes. In order to examine the effect of sample size, we split counties into three groups according to their positive number of reports in the survey: small sizes (less than 15); median sizes (between 15 and 30); large sizes (larger than 30). These summary measures for the comparison are given in Table 1 and 2.

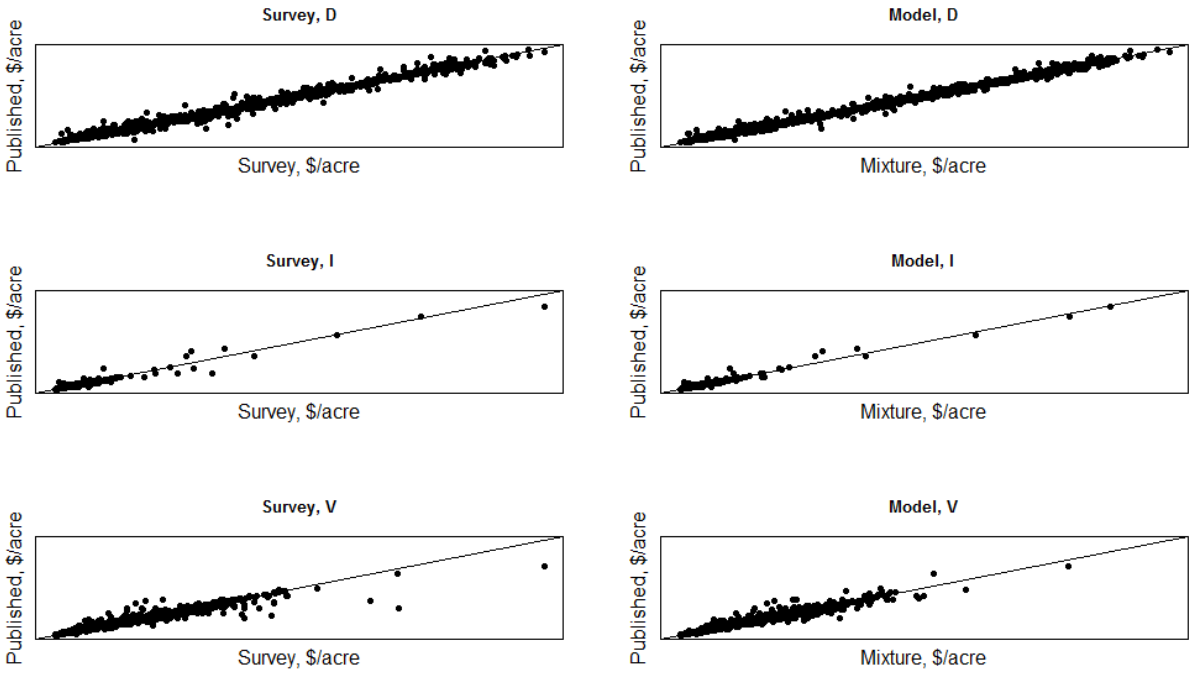

Figure 2: The plots of county-level published estimates for cash rental rates versus survey direct estimates (1st column of the panel) and the model point estimates (2nd column of the panel) for each land type (represented by each row).

To demonstrate the gain in reliability based on the model relative to the survey, we compare the posterior

Table 1: AARD based on nationwide model and survey estimates

Sample Sizes	D		I		V	
	DE	ME	DE	ME	DE	ME
≤ 15	0.0908	**0.0816**	0.0629	**0.0617**	0.0828	**0.0781**
(15,30]	0.0527	**0.0441**	0.0416	**0.0380**	0.0512	**0.0465**
>30	0.0262	**0.0250**	0.0251	**0.0248**	0.0457	**0.0412**

Table 2: ASRD based nationwide model and survey estimates

Sample Sizes	D		I		V	
	DE	ME	DE	ME	DE	ME
≤ 15	0.0351	**0.0224**	0.0135	**0.0129**	0.0233	**0.0151**
(15,30]	0.0084	**0.0063**	0.0054	**0.0048**	0.0102	**0.0065**
> 15	0.0023	**0.0014**	0.0024	**0.0016**	0.0045	**0.0036**

coefficients of variation (CVs) from the model to the survey CVs. The CVs,

$$\text{CV} = 100 \times \frac{\text{SD}_i^m}{\hat{\theta}_i^m}, \tag{3}$$

are calculated, where $m = \{\text{DE, ME}\}$ and SD is the corresponding posterior standard deviation of θ_i^{ME} from the model and the standard error of survey estimates, respectively. Figure 3 shows the CVs comparisons among three groups of counties based on sample sizes. The three scatter plots in the upper row show the CVs comparison between model and survey in small, median, and large sample sizes counties, respectively, for the non-irrigated land (D). The plots in the second row are for the irrigated land and the last row are for the pasture land. The posterior CVs have a greater reduction compared with the CVs of the survey estimates. The median CVs from the model are smaller than the CVs from the survey, for all counties. As expected, the CVs are smaller when sample sizes increase. The results demonstrate the tendency of the small area model to improve the reliability of estimates when compared to the reliability of survey estimates, especially in counties with small sample sizes, that is, counties with very large CVs.

5. Conclusion

We have proposed a two-component mixture HB model with the power prior to obtain the estimates of county-level cash rental rates for US. This is a two-component mixture model that accounts directly for

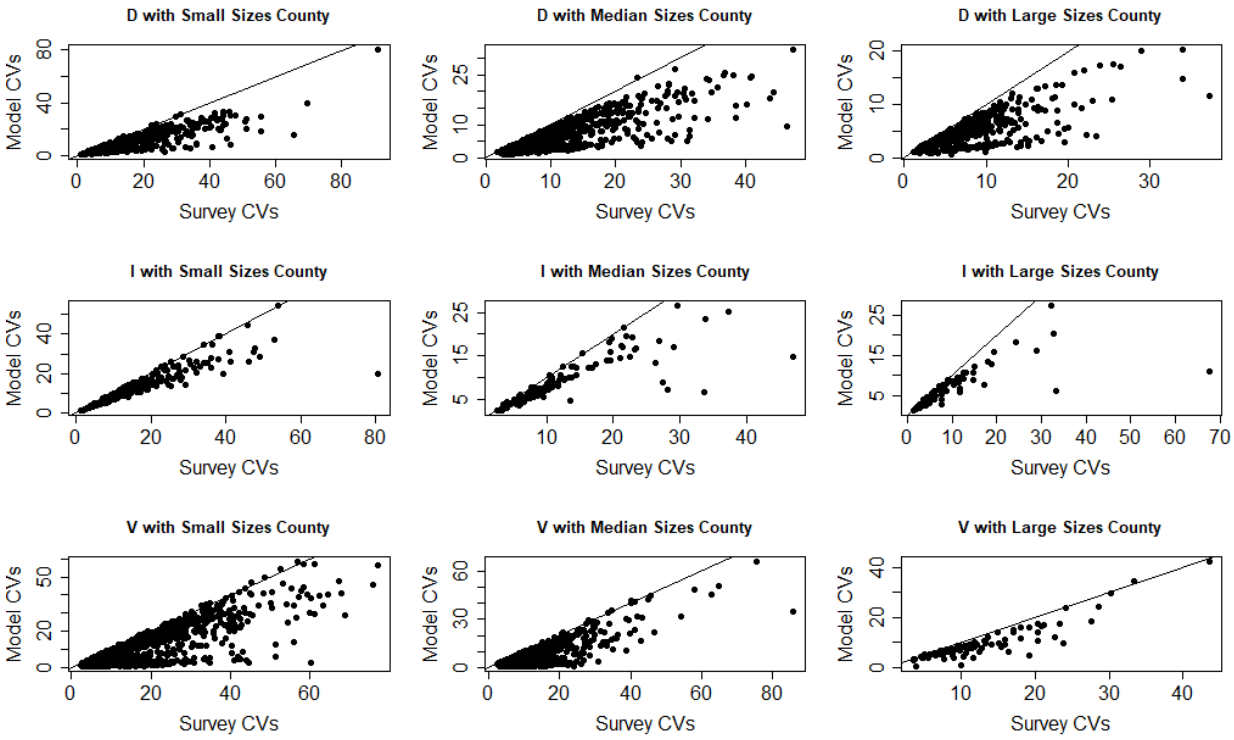

Figure 3: The scatter plots of Model CVs v.s. Survey CVs among small, median, and large size counties by three land types (D, I, V)

outliers and it uses the power prior to input how much of the past data one wants to use. Instead of using all the past data, as is suggested by the two published papers (Berg et al. (2014) and Erciulescu et al. (2019)) on this topic, we model the percentage of the usage of the past data and it varies by the regions in US. In addition, the model avoids correlations between two years by using the power prior. Therefore, the new proposed area-level model is practical in terms of computation time when compared with the unit-level model, which is a key factor for the evaluation of model for the production in government agencies.

We have also included a positivity constraint for the cash rent model. This has been a long-term problem associated with small sample sizes and large sample variances. Without the positivity constraint, it is possible to generate negative estimates, which do not make sense for the cash rental rates. We are able to overcome this problem and propose the computation method with the positivity constraint.

In the case study of 2016 and 2017 CRSs, we show the comparisons between model estimates and survey estimates. First, the scatter plots show that the HB model adjusts for outliers, closer to the published estimates than the survey estimates. In addition, the statistics of AARD and ASRD values show that the model provides estimates closer to the published values than those from the survey. Moreover, the associated measures of uncertainty (CVs) from models are significantly smaller than the CVs of the survey estimates. The model can reduce the CVs while borrowing strength from auxiliary information and all counties within one region. Therefore, the performance of the model illustrates significant improvement of nationwide county-level estimates of cash rental rates for all three land types in accuracy, precision, and reliability.

Ongoing and future research related to the model involve the investigation of different auxiliary information. The auxiliary information considered here is the nationwide data sources available at the county level. Future efforts will be on searching and applying other useful data sources to strengthen the model. Tax data, farm intensity data and other census data are available at the county level in specific states. Variable selections can be investigated for different states and regions by the three land types.

References

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

Berg, E., Cecere, W., and Ghosh, M. (2014). Small Area Estimation for County-Level Farmland Cash Rental Rates. *Journal of Survey Statistics and Methodology*, 2(1):1–37.

Chakraborty, A., Datta, G. S., and Mandal, A. (2019). Robust hierarchical bayes small area estimation for the nested error linear regression model. *International Statistical Review*, 87(S1):S158–S176.

Chen, L., Cruze, N. B., and Young, L. J. (2022a). Model-based estimates for farm labor quantities. *Stats*, 5(3):738–754.

Chen, L., Nandram, B., and Cruze, N. B. (2022b). Hierarchical bayesian model with inequality constraints for us county estimates. *Journal of Official Statistics*, 38(3):709–732.

Chen, M.-H. and Ibrahim, J. G. (2000). Power prior distributions for regression models. *Statistical Science*, 15(1):46 – 60.

Erciulescu, A., Berg, E., Cecere, W., and Ghosh, M. (2019). Bivariate hierarchical bayesian model for estimating cropland cash rental rates at the county level. *Survey methodology*, 45(2):199–216.

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(366a):269–277.

Gershunskaya, J. and Lahiri, P. (2017). Robust empirical best small area finite population mean estimation using a mixture model. *Calcutta Statistical Association Bulletin*, 69(2):183–204.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *IN BAJ. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Eds., Bayesian Statistics*, 4:169–193.

Goyal, S., Datta, G. S., and Mandal, A. (2021). A hierarchical bayes unit-level small area estimation model for normal mixture populations. *Sankhya B*, 83(1):215–241.

Nandram, B., Erciulescu, A., Cruze, N. B., and Chen, L. (2022). Hierarchical bayesian model with inequality constraints for us county estimates. *Research Report RDD-22-02, National Agricultural Statistics Service, USDA*.

National Academies of Sciences, Engineering, and Medicine (2018). *Improving Crop Estimates by Integrating Multiple Data Sources*. The National Academies Press.

Rao, J. N. K. and Molina, I. (2015). *Small Area Estimation*. 2015 John Wiley and Sons, Inc.

APPENDIX A: Integrating out the θ_{ti} from the Joint Posterior Density with Positivity Constraint

We describe a generic problem that can be easily used to obtained the joint density of the $\hat{\theta}_{ti}$. The generic problem is

$$\hat{\theta} \mid \theta \sim \text{Normal}(\theta, \sigma^2), \theta \sim \text{Normal}(\underset{\sim}{x}'\underset{\sim}{\beta}, \delta^2), \theta > 0,$$

where we want to find the marginal distribution of $\hat{\theta}$.

Clearly,

$$p(\hat{\theta}) = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\hat{\theta}-\theta)^2} \times \frac{\frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{1}{2\delta^2}(\theta - \underset{\sim}{x}'\underset{\sim}{\beta})^2}}{\Phi(\frac{\underset{\sim}{x}'\underset{\sim}{\beta}}{\sqrt{\delta^2}})} d\theta. \tag{A.1}$$

Letting $\omega = \frac{\delta^2}{\delta^2+\sigma^2}$, we have

$$\frac{1}{\sigma^2}(\hat{\theta}-\theta)^2 + \frac{1}{\delta^2}(\theta - \underset{\sim}{x}'\underset{\sim}{\beta})^2 = \frac{1}{(1-\omega)\delta^2}\{\theta - (\omega\hat{\theta} + (1-\omega)\underset{\sim}{x}'\underset{\sim}{\beta})\}^2 + \frac{\omega}{\delta^2}(\hat{\theta} - \underset{\sim}{x}'\underset{\sim}{\beta})^2$$

$$p(\hat{\theta}) = \sqrt{\frac{1-\omega}{2\pi\sigma^2}} e^{-\frac{\omega}{2\delta^2}(\hat{\theta} - \underset{\sim}{x}'\underset{\sim}{\beta})^2} \frac{1}{\Phi(\frac{\underset{\sim}{x}'\underset{\sim}{\beta}}{\sqrt{\delta^2}})} \times \int_0^\infty \frac{1}{\sqrt{2\pi(1-\omega)\delta^2}} e^{-\frac{1}{2(1-\omega)\delta^2}\{\theta - (\omega\hat{\theta} + (1-\omega)\underset{\sim}{x}'\underset{\sim}{\beta})\}^2} d\theta.$$

Finally, using a transformation in the integral to a standard normal, it is easy to show that

$$p(\hat{\theta}) = \sqrt{\frac{\omega}{2\pi\delta^2}} e^{-\frac{\omega}{2\delta^2}(\hat{\theta} - \underset{\sim}{x}'\underset{\sim}{\beta})^2} \frac{\Phi\{\frac{\omega\hat{\theta} + (1-\omega)\underset{\sim}{x}'\underset{\sim}{\beta}}{\sqrt{(1-\omega)\delta^2}}\}}{\Phi(\frac{\underset{\sim}{x}'\underset{\sim}{\beta}}{\sqrt{\delta^2}})}, -\infty < \hat{\theta} < \infty. \tag{A.2}$$

It is interesting that there are no restrictions on $\hat{\theta}$ because it is still in $(-\infty, \infty)$. However, the restriction on θ has changed its distribution quite a bit. Doing the algebra again without the constraint, $\theta > 0$ (same as taking the ratio of the two probabilities equal 1 in (A.2)). That is, without any restriction on θ, we have

$$\hat{\theta} \sim \text{Normal}(\underset{\sim}{x}'\underset{\sim}{\beta}, \delta^2 + \sigma^2),$$

as it must be. Actually, this is the distribution in (A.2) but it is adjusted by the ratio of the two normalization constants. In fact, the ratio of the normalization constants goes to 1 as δ^2 goes to 0. That is, in the limit the final density is sensible because θ becomes a point mass at 0.