Small Area Estimation for Cropland Cash Rental Rates

Will Cecere[1], Emily Berg[1], Malay Ghosh[2]
[1]National Agricultural Statistics Service, 3251 Old Lee Highway, Fairfax, VA 22031
[2]Department of Statistics, University of Florida, Gainesville, FL

Abstract

The National Agricultural Statistics Service (NASS) is responsible for estimating average cash rental rates at the county level on an annual basis. A cash rent is land rented on a per acre basis for cash only. Estimates of cash rental rates are useful to farmers in determining rental agreements, economists in studying research questions, and policy makers in computing payment rates for the Conservation Reserve Program. NASS collects data on cash rents using an annual Cash Rent Survey. Because realized sample sizes at the county level are often too small to support reliable direct estimators, mixed model predictors are investigated. A bivariate model is specified to obtain predictors of 2010 cash rental rates for non-irrigated cropland using data from the 2009 Cash Rent Survey and auxiliary variables from external sources such as the 2007 Census of Agriculture. Bayesian methods are used for inference, and results are presented for Iowa, Kansas, and Texas. Incorporating the 2009 survey data through a bivariate model leads to predictors with smaller mean squared errors than predictors based on a univariate model.

Key Words: Small area estimation, bivariate mixed model, benchmarking

1. Introduction

The National Agricultural Statistics Service (NASS) conducts hundreds of surveys each year to obtain estimates related to diverse aspects of US agriculture. NASS's large scale surveys produce reliable estimates of agricultural characteristics at national and state levels. Estimation for small domains, such as counties, is more difficult due to small realized sample sizes.

The focus of our study is estimation of average cash rental rates for non-irrigated cropland at the county level. In a cash rental agreement, a tenant rents cropland or pastureland from a landowner in units of dollars per acre. Cash rental agreements differ from share-rental agreements, which involve payments in terms of a share of the produced goods. The tenant in a cash rental agreement is typically responsible for all management decisions, acquires all of the produced goods, and assumes the risk of achieving lower than expected production due to disease or poor weather.

NASS estimates of county-level cash rental rates serve many purposes. Farmers use the estimates of local cash rental rates for guidance in determining appropriate rental agreements (Dhuyvetter and Kastens, 2009). Agronomists use the estimates to study research questions related to the interplay between cash rental rates and other economic characteristics such as commodity prices and fuel costs (Woodard et al., 2010) . NASS published county-level cash rent estimates have immediate implications for the Conservation Reserve Program, a policy that aims to protect natural resources by providing rental payments to agricultural landowners who choose to conserve their land. Because of the role of cash rental rates in the Conservation Reserve Program, the 2008 Farm Bill required NASS to provide annual estimates of cash rental rates at the county level for three land use categories.

To satisfy the requirements of the 2008 Farm Bill, NASS implemented an annual Cash Rent Survey. A concern is that direct estimators of county means from the Cash Rent Surveys are unstable due to small realized sample sizes. We investigate the use of mixed models (Rao, 2003) to stabilize the estimators of county-level cash rental rates.

Ultimately, estimates are desired for the three land use categories (non-irrigated cropland, irrigated cropland, and pastureland) for counties with at least 20,000 acres of cropland or pastureland. The present study focuses on non-irrigated cropland in Iowa, Kansas, and Texas. The data our analysis include the responses to the 2009 and 2010 Cash Rent Surveys as well as external sources of auxiliary information.

We specify a bivariate, unit-level model to incorporate the correlation between the 2009 and 2010 cash rental rates and use Bayesian methods for inference. Incorporating information from 2009 leads to predictors with smaller mean squared errors than predictors based on a univariate model that only uses the 2010 data. Datta et al. (1998) examine hierarchical Bayes (HB) bivariate models for the county crop acreage data of Battese et al. (1988). We extend the Datta et al. (1998) model to account for non-constant error variances and sets of observations that are not observed in both 2009 and 2010.

NASS estimates state-level cash rental rates using a combination of data from a national survey called the June Area Survey and the Cash Rent Survey. The state-level cash rent estimates are published before county level estimation from the Cash Rent Survey is complete. To maintain internal consistency, it is important that appropriately weighted sums of county predictors are equal to the previously published state estimates. Ghosh et al. (2010) review benchmarking procedures in a Bayesian framework. We use the benchmarking procedure proposed by Ghosh and Steorts (2011) to ensure that the county predictors preserve the previously published state estimates.

In Section 2, we discuss the various data sources available for prediction, which include the responses to the 2009 and 2010 Cash Rent Surveys and covariates from sources external to the Cash Rent Surveys. We describe how we use a bivariate hierarchical Bayes model for inference in Section 3. In Section 4, we summarize results for non-irrigated cropland in Iowa, Kansas, and Texas. We conclude in Section 5 with a summary and areas for future work.

2. Data for Modeling Non-irrigated Cropland Cash Rental Rates

2.1 NASS Cash Rent Survey

NASS implemented an annual Cash Rent Survey in response to the 2008 Farm Bill. The specific objective of the Cash Rent Survey is to obtain county-level estimates of average cash rental rates in three land use categories: non-irrigated cropland, irrigated cropland, and pastureland. The data for our study are from the 2009 and 2010 Cash Rent Surveys.

2.1.1 NASS Cash Rent Survey Sampling Design and Questionnaire

The 2009 and 2010 Cash Rent Surveys used a stratified sampling design. To define the stratification, nine groups were formed on the basis of the dollars rented that an operation reported on previous surveys and censuses. The sampling strata are the intersections of the nine groups and agricultural statistics districts. An agricultural statistics district is a group of contiguous counties within a state that are thought to have similar agricultural characteristics. The sampling fractions within strata are defined so that operations with higher dollars rented on previous surveys and censuses have greater probabilities of selection. The same sample was used for the 2009 and 2010 Cash Rent Surveys, which had a national sample size of approximately 224,000 operations.

The questionnaires for the 2009 and 2010 Cash Rent Surveys consist of three sections. The first set of questions determines how many acres the operation owned, rented or leased from others, and rented to others. The second section determines the total acres operated, which is obtained by subtracting the acres rented to others from the sum of the acres owned and the acres rented from others. The third section of the questionnaire is the most important section for our purposes. The third section asks farmers to report the acres rented for cash in the three land use categories, irrigated cropland, non-irrigated cropland, and pastureland. For each land use category, the farmer is asked to report the total dollars paid or the rental rate in dollars per acre.

A direct survey estimator for a particular land use category is a ratio of a weighted sum of the dollars rented to a weighted sum of acres rented. The weight associated with a respondent is the population size of the stratum containing the respondent divided by the number of responding units in that stratum.

2.1.1 Relationships between 2009 and 2010 Non-irrigated Cropland Cash Rents

The direct estimates for non-irrigated cropland in Iowa for 2010 are plotted against the corresponding direct estimates for 2009 in Figure 1. The correlation between the 2009 and 2010 direct estimates of cash rental rates for non-irrigated cropland in Iowa is 0.8. The county in Iowa with the largest direct estimate for 2009 departs from the linear trend. The correlations between the 2009 and 2010 direct estimates for Kansas and Texas are 0.7 and 0.6, respectively.

To measure the correlation between the reported 2009 and 2010 cash rental rates at the unit (record) level, we computed differences between unit-level cash rental rates for non-irrigated cropland and the mean for a county. Only individuals that reported a cash rental rate for non-irrigated cropland in both years were used to compute the differences. The difference for year t is $y_{ijt} - \bar{y}_{i.t}$, where y_{ijt} is the cash rent per acre for non-irrigated cropland reported by operator j in county i and year t and $\bar{y}_{i.t}$ is the sample average of the y_{ijt} in county i that reported a non-irrigated cropland cash rental rate in both 2009 and 2010. The residuals for Kansas are plotted in Figure 2. The residuals for 2009 and 2010 for Kansas are linearly related, and the correlation between the residuals for 2009 and the residuals for 2010 is 0.7. The extreme values in Figure 2 reflect the high variability among the non-irrigated cropland cash rental rates within a county in Kansas.

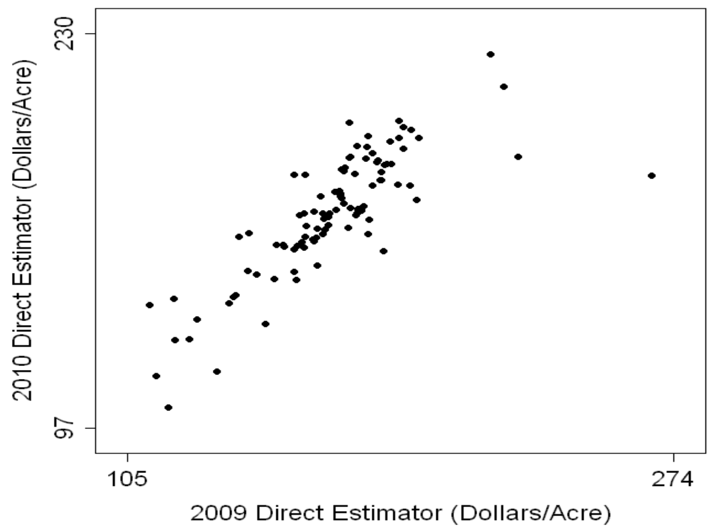

Figure 1: Direct estimates of cash rental rates for non-irrigated cropland in Iowa counties from the 2009 (x-axis) and 2010 (y-axis) Cash Rent Surveys

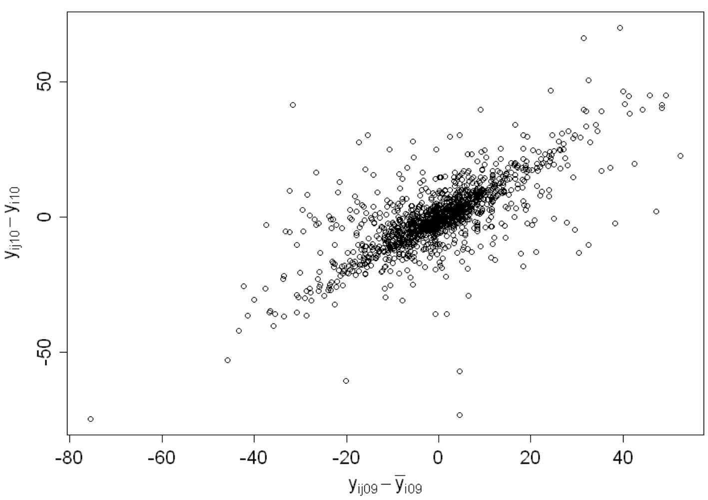

Figure 2: Deviations of unit-level cash rental rates from county means for 2009 (x-axis) and 2010 (y-axis) for units reporting non-irrigated cash rental rates in both years.

Table 1: Correlations between non-irrigated cropland cash rental rates for 2009 and 2010

State	Correlations between 2009 and 2010 direct estimators of average cash rental rates at the county level	Correlations between $y_{ij09} - \bar{y}_{i.09}$ and $y_{ij10} - \bar{y}_{i.10}$ for units reporting non-irrigated rental rates in both years
Iowa	0.81	0.69
Kansas	0.85	0.74
Texas	0.62	0.69

2.2 Auxiliary Information

In an effort to improve the precision of the county-level cash rent estimates, auxiliary variables were desired that would explain both the variability among the county estimates as well as the variability among units within a county. Auxiliary information for modeling cash rental rates is available from several sources external to the Cash Rent Survey. The potential covariates are related to land quality, the value of the commodity sold, and other farm characteristics. Table 2 lists the set of available covariates and indicates whether each covariate is recorded at the county level or the unit level. The covariates are described in more detail in the subsequent subsections.

Table 2: Potential Covariates for Modeling Non-irrigated Cropland Cash Rental Rates

National Commodity Crop Productivity Indexes (NCCPI) at the county level : NCCPI-corn, NCCPI-cotton, NCCPI-wheat, max-NCCPI
Average corn yields at the county level from 2005 – 2009 (Iowa only)
Non-irrigated yield index at the county level (Kansas only)
Total value of production for a county based on the 2007 Census of Agriculture
Farm type (unit-level) – 2 categories: (1) grains/oilseeds, (2) other
Expected sales for an operation recorded on the NASS list frame
Acres rented for non-irrigated cropland recorded on the NASS Cash Rent Survey

2.1.1 Land Quality

One group of potential covariates related to land quality is the National Commodity Crop Productivity Index (NCCPI), which is developed and maintained by the Natural Resource Conservation Service. The NCCPI consists of three different indexes called NCCPI-corn, NCCPI-cotton, and NCCPI-wheat, which reflect the quality of the soil for growing non-irrigated crops in three different climate conditions. The indexes are constructed at the level of a "map unit," a subset of a county. (User Guide National Commodity Crop Productivity Index (NCCPI). Version 1.0, 2008. ftp://ftp-fc.sc.egov/usda.gov/NSCC/ NCCPI_user_guide.pdf.) The county-level indexes used as covariates are weighted averages of the map-unit values, where the weights are the acres of cropland covered by a map unit. The NRCS also produces a summary index called max-NCCPI. The max-NCCPI is obtained by first taking the maximum of the three commodity-specific indexes for each map unit and then computing a weighted average of the maxima across map units in a particular county. The NCCPI provides four potential covariates: the three commodity specific indexes and the max-NCCPI.

Another measure of the quality of the land in a county is the realized crop yield. NASS publishes estimates of crop yields for a variety of crops in many counties. Because the availability of yield data varies by state, the yield covariates differ across states. For Iowa, a yield covariate for county i is a trimmed mean of published corn yields for county i from 2005 through 2009, where the largest and smallest published yields are omitted from the average. All counties in Iowa have a corn yield estimate for at least one of the years, and years for which a yield estimate is missing for a county are excluded from the average for that county. Because Kansas is more agriculturally diverse than Iowa, no single crop is published in at least one year between 2005 and 2009 for all counties of interest. To obtain a covariate that is measured for all counties, we constructed a non-irrigated yield index for Kansas. To construct the yield index, we first averaged the published yields for corn, wheat, and sorghum using the method described for the Iowa corn yields. The average yields were then standardized to have mean zero and variance one. The non-irrigated yield index for a county is defined as the largest of the three standardized yields. Counties without published yields are excluded from the standardization and maximization. Table 3 illustrates the construction of the non-irrigated yield index. For Texas, 38 counties with reported values for non-irrigated cash rental rates in 2009 or 2010 have no published yields for any of the years 2005 through 2009. Because of the insufficient yield data in Texas, we did not use a yield covariate for Texas.

Table 3: Kansas Non-irrigated Yield Index

County	Standardized Yield			Index
	Corn	Sorghum	Wheat	
Atchison	1.025	2.971	1.292	2.971
Cherokee	-0.059	NA	0.397	0.397

2.2.2 Value of the Commodity Sold

Two measures of the value of the goods that an operation produces and sells are included in the list of potential covariates: the total value of production (TVP) and the expected sales. The TVP is a county-level covariate obtained from the 2007 Census of Agriculture. The expected sales is a unit-level covariate obtained from the NASS list frame.

2.2.3 Other Farm Characteristics

Two farm characteristics were considered as potential covariates: the acres rented for non-irrigated cropland and the farm type. The acres rented for non-irrigated cropland associated with unit (ij) in year t is the reported acres rented on the current year's Cash Rent Survey for year t. Farms are partitioned into 17 farm types on the NASS list frame. To define a covariate to use for modeling non-irrigated cash rental rates, the farm types were aggregated into two groups. The first group consists of farms that produce grains, oilseeds and dried beans. All other farm types were combined to form the second group.

3. Bivariate Hierarchicial Bayes Model

The correlation between the 2009 and 2010 cash rental rates observed in Section 2.1.1 suggests that using the information in the data from 2009 has the potential to improve the predictions for 2010. A bivariate hierarchical model is specified as a way to incorporate the data for both years. Let a_{ijt} and y_{ijt} be the acres and dollars per acre, respectively, rented by operator j in county i and year t $(t = 09, 10)$, and let $\boldsymbol{x}_{ijt}$ be the associated column vector of auxiliary variables with dimension p_t. For covariates that are constant across years and individuals, $\boldsymbol{x}_{ijt} = \boldsymbol{x}_{i109}$. Let $w_{ijt} = a_{ijt}N_{g(ijt)}n_{g(ijt)}^{-1}$, where $N_{g(ijt)}$ and $n_{g(ijt)}$ are the population size and number of respondents, respectively, in year t for the stratum g that contains unit (ij).

To specify the model, we divide the respondents into three sets:

- Set 1 consists of units (ij) that reported a non-irrigated cash rental rate in both 2009 and 2010.
- Set 2 consists of units (ij) that only reported a non-irrigated cash rental rate in 2009.
- Set 3 consists of units (ij) that only reported a non-irrigated cash rental rate in 2010.

We assume that observations in set 1 satisfy the bivariate model,

$$\begin{pmatrix} y_{ij09} \\ y_{ij10} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}'_{ij09}\boldsymbol{\beta}_{09} + v_{i09} + e_{ij09} \\ \boldsymbol{x}'_{ij10}\boldsymbol{\beta}_{10} + v_{i10} + e_{ij10} \end{pmatrix}, \tag{1}$$

where

$$\begin{pmatrix} e_{ij09} \\ e_{ij10} \end{pmatrix} \sim N\left(\boldsymbol{0}, \boldsymbol{D}_{wij}^{-0.5}\boldsymbol{\Sigma}_{ee}\boldsymbol{D}_{wij}^{-0.5}\right), \tag{2}$$

$\boldsymbol{D}_{wij} = diag(w_{ij09}, w_{ij10})$, and

$$\begin{pmatrix} v_{i09} \\ v_{i10} \end{pmatrix} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_{vv}). \tag{3}$$

We denote the diagonal elements of $\boldsymbol{\Sigma}_{ee}$ corresponding to 2009 and 2010 by σ_{ee09} and σ_{ee10}, respectively. For units (ij) in set 2, we assume,

$$y_{ij09} = \boldsymbol{x}'_{ij09}\boldsymbol{\beta}_{09} + v_{i09} + e^*_{ij09}, \tag{4}$$

where $e^*_{ij09} \sim N(0, w_{ij09}^{-1}\tau^2_{e09})$. The regression coefficient, $\boldsymbol{\beta}_{09}$, and the county effect, v_{i09}, in the models for set 1 and set 2 are the same. The variance of e^*_{ij09} differs from the diagonal element of $\boldsymbol{\Sigma}_{ee}$ associated with 2009. The model for set 3 has the same form as the model for set 2 except the subscripts 09 are replaced by 10. Specifically, for units (ij) in set 3, we assume,

$$y_{ij10} = \boldsymbol{x}'_{ij10}\boldsymbol{\beta}_{10} + v_{i10} + e^*_{ij10}, \tag{5}$$

where $e^*_{ij10} \sim N(0, w_{ij10}^{-1}\tau^2_{e10})$.

The quantity of interest for 2010 is,

$$\theta_{i10} = \overline{\boldsymbol{x}}'_{wi10}\boldsymbol{\beta}_{10} + v_{i10}, \tag{6}$$

where $\overline{\boldsymbol{x}}'_{wi10} = \left(\sum_{j=1}^{n_{i10}} w_{ij10}\right)^{-1}\left(\sum_{j=1}^{n_{i10}} w_{ij10}\boldsymbol{x}_{ij10}\right)$ and n_{i10} is the total number of respondents to the 2010 Cash Rent Survey who reported a non-irrigated cropland cash rental rate. (Equivalently, n_{i10} is the sum of the number of units in set 1 and in set 3.)

The variances of the unit-level errors, e_{ijt} and e^*_{ijt} are assumed to be inversely proportional to the weight, w_{ijt}, for two reasons. First, if $\boldsymbol{\Sigma}_{ee}$ is diagonal, and $\tau^2_{et} = \sigma_{eet}$, then an EBLUP for county i in year t is a convex combination of the direct estimator and an estimator of $\overline{\boldsymbol{x}}'_{wit}\boldsymbol{\beta}_t$, where the weight assigned to the direct estimator in the convex combination increases as the county sample size increases. Second, the variances of residuals from preliminary analyses decrease as the acres increase.

Diffuse, proper priors are specified for the unknown regression coefficients and variances. Specifically, $\boldsymbol{\beta}_t \sim N(0, 10^6\boldsymbol{I})$, and $\tau^2_{et} \sim \text{inverse} - \text{gamma}(0.001, 0.001)$. The covariance matrices, $\boldsymbol{\Sigma}_{ee}$ and $\boldsymbol{\Sigma}_{vv}$, have inverse-Wishart prior distributions with shape parameter 0.01 and a diagonal scale matrix with diagonal elements 0.001. The parameterizations for the inverse-gamma and inverse-Wishart distributions are from Gelman et al. (2009).

3.1 Gibbs Sampling and Posteriors

We use Gibbs sampling to obtain a Monte Carlo approximation to the posterior distribution. An initial analysis of BGR statistics (Gelman et al., 2009) based on three MCMC chains, each with 20,000 iterations, indicated that 1000 iterations is sufficient for burn-in. The analyses in Section 4 are based on one chain of length 20,000 for each of the three states, Iowa, Kansas and Texas, where the first 1000 iterations are discarded for burn-in. By the choices of the likelihood and the priors, the full conditional distributions are known distributions. (See the Appendix for details.) The Bayes predictor of θ_{i10} for squared error loss is $\hat{\theta}_{i10}^{B} = E[\theta_{i10}|\boldsymbol{y}]$, where $\boldsymbol{y}$ is the vector of observed non-irrigated cash rental rates.

3.2 Two-stage benchmarking

NASS obtains estimates of cash rental rates at the state level using data from a national survey that is conducted in June in addition to the Cash Rent Survey. The state estimates are published before the county-level data from the Cash Rent Survey are fully processed. NASS also establishes estimates of cash rental rates for agricultural statistics districts, groups of spatially contiguous counties within a state. To retain internal consistency, it is important that appropriately weighted sums of county estimates equal the district estimates and appropriately weighted sums of district estimates equal the previously published state estimate. The benchmarking restrictions for a single time-point are,

$$\sum_{i \epsilon d_k} w_i \hat{\theta}_i = \hat{\lambda}_k, \tag{7}$$

and

$$\sum_{k=1}^{D} \eta_k \hat{\lambda}_k = \theta_{pub}, \tag{8}$$

where $w_i = (\sum_{i\epsilon d_k} z_i)^{-1} z_i$, $\eta_k = (\sum_{k=1}^{D} \sum_{i\epsilon d_k} z_i)^{-1} \sum_{i\epsilon d_k} z_i$, z_i is a direct estimate of the acres rented in county i, d_k is a set of indexes for the counties in district k, $\hat{\theta}_i$ is the final estimate of the average cash rental rate for county i, $\hat{\lambda}_k$ is the final estimate of the average cash rental rate for district k, and θ_{pub} is the published estimate of the state-level cash rent per acre. The index for the year is suppressed in (7) and (8) for simplicity. The direct estimators of the acres rented at the county and district levels are treated as fixed for our analysis.

We use the two-stage benchmarking procedure proposed by Ghosh and Steorts (2011) to define benchmarked estimates. The benchmarked estimates minimize the quadratic form,

$$g(\boldsymbol{c},\boldsymbol{d}) = \sum_{k=1}^{D} \sum_{i\epsilon d_k} \zeta_i \left(\hat{\theta}_i^B - c_i\right)^2 + \sum_{k=1}^{D} \rho_k (\hat{\theta}_{k,w}^B - d_k)^2 \tag{9}$$

subject to the constraints in (7) and (8), where $\boldsymbol{c} = (c_1, \ldots, c_m)$, $\boldsymbol{d} = (d_1, \ldots, d_K)$, $\hat{\theta}_{k,w} = \sum_{i\epsilon d_k} w_i \hat{\theta}_i^B$, and ρ_k and ζ_i are constants selected by the analyst. We choose $\zeta_i = w_i$ and $\rho_k = \eta_k$, which gives the benchmarked estimates,

$$\hat{\theta}_i = \hat{\theta}_i^B + \hat{\lambda}_{k(i)} - \hat{\theta}_{k(i),w}^B, \tag{10}$$

and

$$\hat{\lambda}_{k(i)} = \hat{\theta}^{B}_{k(i),w} + \frac{(\theta_{pub} - \hat{\theta}^{B}_{w})\eta_{k(i)}(1+\eta_{k(i)})^{-1}}{\sum_{i\epsilon d_{k(i)}} \eta^{2}_{k(i)}(1+\eta_{k(i)})^{-1}}, \tag{11}$$

for county i and district $k(i)$, respectively, where $k(i)$ is the district containing county i. In (11) , $\hat{\theta}^{B}_{w} = \sum_{k=1}^{D} \eta_k \hat{\theta}^{B}_{k,w}$. Each of the benchmarked estimates in (10) and (11) is a sum of the hierarchical Bayes predictor and an adjustment term. If the hierarchical Bayes predictor for the state is larger (smaller) than the previously published state total, then the adjustment is negative (positive), and the benchmarked county and district estimates are smaller (larger) than the hierarchical Bayes predictors. The posterior mean squared error of the benchmarked predictor for year t is the sum of the posterior variance of θ_{it} and the squared difference between the Bayes predictor and the benchmarked predictor. See (You et al., 2002) for a derivation of the posterior MSE of a benchmarked predictor.

4. Results for Non-irrigated Cropland in Iowa, Kansas, and Texas

The model of Section 3 was fit to the non-irrigated cropland cash rental rates reported on the 2009 and 2010 Cash Rent Surveys for Iowa, Kansas, and Texas. These three states were chosen to reflect a range of challenges. All counties in Iowa have estimates for corn yields, and cash renting is a relatively common way to rent non-irrigated cropland. Kansas is more agriculturally diverse than Iowa. According to agricultural specialists at NASS, share-renting is a more common way to rent land than cash renting in many parts of Texas, which may explain why realized sample sizes for some Texas counties are as small as zero or one report.

4.1 Covariate Selection

The potential covariates for Iowa, Kansas, and Texas are listed in Table 2. For each state, the covariates include four variables related to the NCCPI, the total value of production for a county based on the 2007 Census of Agriculture, the expected sales for an operation recorded on the NASS list frame, the farm type recorded on the NASS list frame, and the and the acres rented for non-irrigated cropland recorded on the NASS Cash Rent Survey.

The covariates for each state were selected according to the following procedure. First, univariate models were fit to the data for 2009 and 2010 separately using maximum likelihood estimation. The univariate model used for covariate selection is of the form,

$$y_{ijt} = \boldsymbol{x}'_{ijt}\boldsymbol{\alpha}_t + v_{it} + \varepsilon_{ijt}, \tag{12}$$

where $\varepsilon_{it} \sim N(0, \sigma^2_{\varepsilon,t})$, and $v_{it} \sim N(0, \sigma^2_{v,t})$. All of the units who reported a non-irrigated cropland cash rental rate in year t were used to fit the univariate model for year t, regardless of whether or not the unit also reported a cash rental rate in year s $(s \neq t)$. The R function lmer in the package nlme was used for maximum likelihood estimation. For each year, step-wise selection using the R function stepAIC was performed using the BIC measure of goodness of fit. The selected covariates are the variables that are in the minimum BIC models for both the 2009 and 2010 univariate models. Table 4 contains the covariates that were selected for Iowa, Kansas, and Texas.

Table 4: Selected Covariates

Iowa:	corn yield, expected sales, non-irrigated acres rented for cash
Kansas:	non-irrigated yield index, expected sales, farm type
Texas:	max-NCCPI, expected sales, farm type

4.2 Estimates of Correlation Parameters

The exploratory plots and correlation estimates in Section 2.1 suggest a substantial correlation between the non-irrigated cropland cash rental rates for 2009 and 2010 at both the unit and county levels. Table 5 contains summaries of the posterior distributions of the correlations in the model from section 3.1. The columns labelled "Median" are the posterior medians of the correlations, and lower and upper endpoints of the 95% credible intervals are the 2.5 and 97.5 percentiles of the posterior distributions of the correlations. Even though the variances of e_{ij09} and e_{ij10} are proportional to the inverses of the weights, the correlation is a constant because the weights cancel in the definition of the correlation.

Table 5: Posterior distributions of correlations between 2009 and 2010

	$Cor\{v_{i09}, v_{i10}\}$		$Cor\{e_{i09}, e_{i10}\}$	
State	Median	95% Credible Interval	Median	95% Credible Interval
Iowa	0.787	[0.656, 0.860]	0.571	[0.548, 0.592]
Kansas	0.909	[0.855, 0.944]	0.727	[0.700, 0.751]
Texas	0.884	[0.832, 0.922]	0.691	[0.666, 0.714]

The posterior medians of the county-level and unit-level correlations exceed 0.75 and 0.57, respectively. The lower endpoints of the 95% credible intervals exceed 0.76 and 0.54 for the unit-level and county-level correlations, respectively. For each state, the correlations at the level of the county are larger than the correlations for individual units. The significant correlations suggest the potential for an efficiency gain relative to a univariate model.

4.3 Comparison of 2010 Predictors for Bivariate and Univariate Models

To demonstrate the gain in efficiency due to the use of the bivariate model relative to a univariate model, we compared the posterior mean squared errors of the predictors from the bivariate model to the posterior mean squared errors of the predictors from a corresponding univariate model. The assumptions of the univariate models are the same as the assumptions of the bivariate models except that the covariance parameters in $\mathbf{\Sigma}_{ee}$ and $\mathbf{\Sigma}_{vv}$ are assumed to equal to zero. To fit the univariate models, we use inverse-gamma prior distributions for σ_{eet} and σ_{vvt} $(t = 09, 10)$.

We define the efficiency of a 2010 predictor for county i based on a bivariate model relative to the efficiency of a 2010 predictor based on a univariate model by,

$$REFF_{i,10} = \frac{MSE_{BV}\left(\hat{\theta}_{i10}^{BV}\right)}{MSE_{UNI}\left(\hat{\theta}_{i10}^{UNI}\right)}, \tag{13}$$

where $MSE_{BV}\left(\hat{\theta}_{i10}^{BV}\right)$ is the posterior MSE of the benchmarked predictor for 2010 based on the bivariate model and $MSE_{UNI}\left(\hat{\theta}_{i10}^{UNI}\right)$ is the posterior MSE based on the corresponding univariate model. Table 6 shows the average relative efficiencies for Iowa, Kansas, and Texas, where the average relative efficiency is defined as the average of $REFF_{i,10}$ across the counties in a state. Because of the significant correlations in the model errors, the posterior MSE from a bivariate model is smaller than the posterior MSE from the corresponding univariate model, and the average relative efficiencies are less than one.

Table 6: Averages of ratios of posterior MSE's from a bivariate model to posterior MSE's from a univariate model.

State	Average Relative Efficiency
Iowa	91.1%
Kansas	90.9%
Texas	85.0%

5. Conclusions and Future Work

We use a bivariate HB model to obtain predictors of county-level cash rental rates for non-irrigated cropland in Iowa, Kansas, and Texas. The model incorporates auxiliary information related to land quality, commodity values, and farm characteristics. Significant correlations exist between the 2009 and 2010 model random effects at both the unit and county level. As a consequence, using the information in the 2009 cash rent estimates reduces the posterior MSE relative to a univariate model.

One area for future work involves extending the procedures developed for non-irrigated cropland to all states and to the other two land use categories of interest to NASS (irrigated cropland, and pastureland). The current approach treats the estimated acres rented as fixed values. Future work may involve some form of modeling or systematic adjustments to reflect errors in the reported and estimated acreages. An analysis of deleted residuals suggests that the distributions of the errors have heavier tails than the assumed normal distribution. One option for future work is to specify a heavy-tailed distribution for the errors that may represent the observed responses more appropriately than the assumed normal distribution. The benchmarking procedure is applied to each time point separately. In the interest of improving the efficiency of the benchmarked predictors, we may investigate a bivariate benchmarking operation that recognizes the correlation between the two time points.

Acknowledgements

The authors would like to thank Angie Considine, Sharyn Lavender, Curt Stock, Scott Shimmin, and Dan Beckler from the National Agriculture Statistics Service. Additional thanks are in order to Bob Dobos from the National Resource Conservation Service and Rich Iovanna from the Farm Service Agency. Without the generous assistance from these individuals, this research would be of significantly less quality.

References

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

Datta, G.S., Day, B., Maiti, T. (1998). “Multivariate Bayesian Small Area Estimation: An Application to Survey and Satellite Data,’’ *Sankhya: The Indian Journal of Statistics*, 60, 344-362.

Datta, GS., Ghosh M., Steorts, R., and Maples, J. (2010). “Bayesian Benchmarking with Applications to Small Area Estimation,” Unpublished technical report.

Dhuyvetter, D., and Kastens, T. (2010). “Kansas Land Values and Cash Rents at the County Level.” http://www/agmanager.info/farmmt/land/county/CountValues &Rents(Sep2010).pdf.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2009). *Bayesian Data Analysis*: *Second Edition*, CRC Press.

Ghosh, M. and Steorts, R. C., “Two Stage Bayesian Benchmarking as Applied to Small Area Estimation,” (2011). To be submitted.

Woodard, S., Paulson, N., Baylis, K., and Woddard, J. (2010). “Spatial Analysis of Illinois Agricultural Cash Rents,” The Selected Works of Kathy Baylis. http://works.bepress.com/kathy_baylis/29.

Rao J.N.K.(2003): *Small Area Estimation*, New Work: John Wiley and Sons.

You, Y., Rao, J.N.K., Dick, P. (2002). “Benchmarking Hierarchicial Bayes Small Area Estimators with Applications in Census Undercoverage Estimation.” *Proceedings of the Survey Methods Section*, Statistical Society of Canada.

Appendix

To specify the full conditional distributions for Gibbs sampling, we introduce some notation. Let $\boldsymbol{\Theta}_\gamma$ be the set of parameters except for the parameter denoted by γ. Let $\boldsymbol{X}_{ij} = \left(\boldsymbol{z}_{ij,09}, \boldsymbol{z}_{ij,10}\right)'$, where $\boldsymbol{z}_{ij,09} = \left(\boldsymbol{x}'_{ij,09}, \boldsymbol{0}'_{p_{10}}\right)'$, and $\boldsymbol{z}_{ij,10} = \left(\boldsymbol{0}'_{p_{09}}, \boldsymbol{x}'_{ij,10}\right)'$. Let $\boldsymbol{y}'_{ij} = \left(y_{ij,09}, y_{ij,10}\right)$. Let A_i be the set of units in county i that are in set 1, $B_{i,09}$ be the set of units in county i that are in set 2, and $B_{i,10}$ be the set of units in county i that are in set 3, where set 1, set 2, and set 3 are defined in Section 3. Full conditionals are as follows.

1. $\boldsymbol{\beta} \quad | \boldsymbol{\Theta}_{\boldsymbol{\beta}},\ \boldsymbol{y} \sim \mathrm{N}\{\boldsymbol{\Sigma}_{\beta\beta}\boldsymbol{r}_{\beta},\ \boldsymbol{\Sigma}_{\beta\beta}\}$, where

$$\boldsymbol{\Sigma}_{\beta\beta} = \left[\sum_{i=1}^{m}\sum_{j\epsilon A_i}\boldsymbol{X}'_{ij}\boldsymbol{D}^{0.5}_{wij}\boldsymbol{\Sigma}^{-1}_{ee}\boldsymbol{D}^{0.5}_{wij}\boldsymbol{X}_{ij} + 10^{-6}\boldsymbol{I}_{p_{09}+p_{10}} + \boldsymbol{\Omega}\right]^{-1},$$

$$\boldsymbol{\Omega} = \mathrm{diag}\left(\sum_{i=1}^{m}\sum_{j\epsilon B_{i09}} w_{ij,09}\boldsymbol{x}_{ij,09}\boldsymbol{x}'_{ij,09}\ \tau^{-2}_{e,09}, \sum_{i=1}^{m}\sum_{j\epsilon B_{i09}} w_{ij,10}\boldsymbol{x}_{ij,10}\boldsymbol{x}'_{ij,10}\ \tau^{-2}_{e,10}\right),$$

$$\boldsymbol{r}_{\beta} = \sum_{i=1}^{m}\sum_{j\epsilon A_i}\boldsymbol{X}'_{ij}\boldsymbol{D}^{0.5}_{wij}\boldsymbol{\Sigma}^{-1}_{ee}\boldsymbol{D}^{0.5}_{wij}(\boldsymbol{y}_{ij} - \boldsymbol{\nu}_i) + \boldsymbol{r}_{\beta 2},$$

and

$$\boldsymbol{r}_{\beta 2} = \begin{pmatrix} \sum_{i=1}^{m} \sum_{j \epsilon B_{i,09}} \tau_{e,09}^{-2}\, w_{ij,09} \boldsymbol{x}_{ij,09} \left(y_{ij,09} - \nu_{i,09} \right) \\ \sum_{i=1}^{m} \sum_{j \epsilon B_{i,10}} \tau_{e,10}^{-2}\, w_{ij10} \boldsymbol{x}_{ij,10} \left(y_{ij,10} - \nu_{i,10} \right) \end{pmatrix}.$$

2. $\boldsymbol{\Sigma}_{ee} \mid \boldsymbol{\Theta}_{\Sigma_{ee}}, \boldsymbol{y} \sim \text{Inverse} - \text{Wishart}(\boldsymbol{A}_e, d_e)$, where $d_e = \sum_{i=1}^{m} |A_i| + \delta_e$, and

$$\boldsymbol{A}_e = \sum_{i=1}^{m} \sum_{j \epsilon A_i} \boldsymbol{D}_{wij}^{0.5} (\boldsymbol{y}_{ij} - \boldsymbol{\nu}_i - \boldsymbol{X}_{ij} \boldsymbol{\beta})(\boldsymbol{y}_{ij} - \boldsymbol{\nu}_i - \boldsymbol{X}_{ij} \boldsymbol{\beta})' \boldsymbol{D}_{wij}^{0.5}$$

3. $\boldsymbol{\Sigma}_{vv} \mid \boldsymbol{\Theta}_{\Sigma_{vv}}, \boldsymbol{y} \sim \text{Inverse} - \text{Wishart}(\boldsymbol{A}_v, d_v)$, where

$$d_e = m + \delta_e,$$

and

$$\boldsymbol{A}_v = \sum_{i=1}^{m} \boldsymbol{\nu}_i \boldsymbol{\nu}_i'.$$

4. $\tau_{e,t}^2 \mid \boldsymbol{\Theta}_{\tau_{e,t}^2}, \boldsymbol{y} \sim$ Inverse-Gamma(a_{et}, d_{et}), where

$$d_{et} = \sum_{i=1}^{m} |B_{i,t}| + \delta_e,$$

and

$$a_{et} = \sum_{i=1}^{m} \sum_{j \epsilon B_{it}} \boldsymbol{D}_{wij} (y_{ij,t} - \nu_{i,t} - \boldsymbol{x}_{ij,t} \boldsymbol{\beta}_t)^2.$$

5. $\boldsymbol{\nu}_i \mid \boldsymbol{\Theta}_{\nu_i}, \boldsymbol{y} \sim N(\boldsymbol{\mu}_{vv}, \boldsymbol{M}_i^{-1})$, where $\boldsymbol{M}_i = (\boldsymbol{\Sigma}_{vv}^{-1} + \boldsymbol{\Sigma}_{ee,wi}^{-1} + \boldsymbol{\Omega}_{ee,wi}^{-1})^{-1}$, $\boldsymbol{\mu}_{vv} = \boldsymbol{M}_i^{-1}(\boldsymbol{r}_{i1} + \boldsymbol{r}_{i2})$, $\boldsymbol{\Sigma}_{ee,wi}^{-1} = \sum_{j \epsilon A_i} \boldsymbol{D}_{wij}^{0.5} \boldsymbol{\Sigma}_{ee}^{-1} \boldsymbol{D}_{wij}^{0.5}$,

$\boldsymbol{\Omega}_{ee,wi}^{-1} = diag\left(\boldsymbol{\tau}_{e,09}^{-2} w_{i.B09}, \boldsymbol{\tau}_{e,10}^{-2} w_{i.B10}\right)$,

$$\boldsymbol{r}_{i1} = \sum_{j \epsilon A_i} \boldsymbol{D}_{wij}^{0.5} \boldsymbol{\Sigma}_{ee}^{-1} \boldsymbol{D}_{wij}^{0.5} (\boldsymbol{y}_{ij} - \boldsymbol{X}_{ij} \boldsymbol{\beta}),$$

$$\boldsymbol{r}_{i2} = \begin{pmatrix} \sum_{j \epsilon B_{i09}} w_{ij,09} (\boldsymbol{y}_{ij,09} - \boldsymbol{X}_{ij,09} \boldsymbol{\beta}_{09})\, \boldsymbol{\tau}_{e,09}^{-2} \\ \sum_{j \epsilon B_{i10}} w_{ij,10} (\boldsymbol{y}_{ij,10} - \boldsymbol{X}_{ij,10} \boldsymbol{\beta}_{10}) \boldsymbol{\tau}_{e,10}^{-2} \end{pmatrix}.$$