Empirical Bayes Methodology for the NASS County Estimation Program

Michael E. Bellow, USDA-NASS
Partha Lahiri, University of Maryland, College Park

1. Introduction

Various government agencies (e.g., the United States Census Bureau, USDA's National Agricultural Statistics Service (NASS), Statistics Canada and the Central Statistical Office of the United Kingdom) are required to produce reliable small-area statistics. A small-area generally refers to a geographical entity (e.g., county) for which limited information is available from the primary source of data. Accurate small-area statistics are needed for regional planning and fund allocation in many government programs and thus their importance cannot be overemphasized. County estimates of crop parameters such as yield are used by farmers, agribusinesses and government agencies for local agricultural decision making.

NASS has been publishing county level crop and livestock inventories since 1917 (see Iwig, 1993). The main source of data used by the agency for commodity estimation has always been its surveys of farmers, ranchers and agribusiness managers who provide requested information on a voluntary, confidential basis. Since surveys designed and conducted at the national and state levels are seldom adequate for obtaining reliable county level estimates, NASS has made extensive use of ancillary data sources such as list sampling frame control data, previous year estimates, earth observing satellite data and census of agriculture data in its county estimation procedures.

The basic county estimation approach used by NASS has remained relatively unchanged over the years. The procedure for estimating totals such as crop acreages and livestock inventories initially involves scaling the survey estimates and other available administrative data at the county level for consistency with official state level estimates. The scaled estimates are composited together (usually with previous year estimates) to produce county level estimates for the current year, which are checked against available administrative data sources that are considered reliable indicators of minimum levels and modified if necessary. Program changes over the years have been driven by advances in data storage and processing technology as well as improvements in sampling frame and sample selection methodology, most notably the introduction of probability based sampling by NASS in the 1950s and 60s and multiple frame (area and list) methods in the 1970s.

NASS field offices (located in 45 states) have the task of producing estimates of various crop and livestock items at the county level within their respective states. Each field office conducts a separate annual County Estimates Survey (CES). Since 2002, multivariate probability proportional to size (MPPS) sampling has been used to select the CES samples of farms, with questionnaires mailed out to the operators and telephone followups done where necessary. Data from other NASS surveys (such as the September and December Quarterly Agricultural Surveys (QAS) and January Cattle Survey) are merged with the CES sample to form a combined data set which is then used to calculate various commodity indications at the county level. Final county estimation usually takes place after the QAS-based state level estimates have been released. The main software tool employed by the field offices is the Database Integrated County Estimates (DICE) system, a client-server type system that processes input indications from all internal and external data sources used. Outputs of the DICE system are used to set final official county estimates subject to certain consistency and confidentiality requirements. For example, county estimates are constrained to sum to official district (subdivision of a state into agricultural regions) and state level totals. The census of agriculture (conducted twice each decade by NASS in years ending in '2' and '7') serves as a useful benchmark for county estimation.

In general, traditional direct methods that utilize only small-area specific survey data are highly unreliable, mainly due to small sample sizes in the areas of interest. In addition, effects of nonsampling errors such as coverage and nonresponse can be severe, and combining data from several different surveys doesn't often resolve this problem satisfactorily. In order to improve on direct estimation, several indirect and model-based methods have been proposed in the literature. These procedures essentially use implicit or explicit models that *borrow strength* from related resources such as administrative and census records and survey data from previous years. Rao (2003) and Jiang and Lahiri (2006) provide comprehensive reviews of different small-area estimation methods and applications.

A number of small-area methods have been proposed over the years for county estimation of crop items. Stasny, Goel, and Rumsey (1991) developed a unit level regression model to produce county regression synthetic estimates of wheat production in Kansas. Stasny, Goel, Cooley and Bohn (1995) proposed a more sophisticated unit level Bayesian mixed effects model (with a simple spatial component) for estimating crop yields. Griffith (1999, 2001) considered a small-area unit level model that allows for both temporal and spatial correlations. Comparisons of different methods are provided by Crouse (2000) and Bellow (2007).

Data from multiple sample surveys (such as the Acreage and Production Survey (APS) and QAS) are used to estimate harvested yield for various crops at the county level. Currently, standard design-based estimates of soybean harvested yield cannot be obtained from the database since survey weights are not available. The main component of the data is the APS, which is also subject to high nonresponse rates and coverage bias. Kott (2009) proposed a model-based direct estimator of crop items such as harvested yield, based on a county specific regression model with heteroscedastic error variances using size variables from the sampling frame. This method (known as *Kott-Busselberg*) is direct since estimation is based on county specific data for the commodity in question. In a previous study, the Kott-Busselberg (KB) estimator of harvested yield exhibited bias when compared with corresponding census yields – this problem has since been rectified by Lahiri (2010). Since the direct variance estimators are unstable for counties with small sample sizes, Kott proposed a smoothed variance estimator using data from all counties in the agricultural district in which the county resides. From this point on, Lahiri's modified version of the Kott-Busselberg estimator will be referred to as the *direct estimator*.

In Section 2, we propose an empirical Bayes (EB) estimator of harvested yield that improves on the direct estimator by combining it with different auxiliary variables (described in Section 3). The EB estimator is a weighted average of the direct estimator and a regression synthetic estimator that incorporates the ancillary information. A parametric bootstrap mean squared error estimator for EB is also described in Section 2.

In Section 4, we discuss results of a study evaluating soybean harvested yield at the county level for seven states in the midwestern US in 2007. The choice of a year when the census of agriculture was conducted enabled different estimators to be compared with corresponding census figures (regarded as a gold standard). Overall, our EB estimator of harvested yield was found to perform better than the competing estimators.

2. Empirical Bayes Estimation

The following basic area-level model for crop harvested yield combines direct estimates and related county specific auxiliary variables (Fay and Herriot, 1979):

$$\text{\underline{Level 1}} \text{ - } \hat{t}_i^{Dir} \mid t_i \sim ind\, N[t_i, D_i],$$
$$\text{\underline{Level 2}} \text{ - } t_i \sim ind\, N[x_i^T \beta, A] \quad (i=1,\dots,m).$$

where:

m = number of counties in the state,

t_i = true harvested yield of crop in county i,

$\hat{t}_i^{Dir}$ = direct estimator of harvested yield in county i,

D_i = sampling variance of direct estimator,

$x_i^T = (x_{1i}, \dots, x_{pi})$ = vector of auxiliary variables ($i = 1, \dots m$),

$\beta = (\beta_1, \dots, \beta_p)$ = vector of regression parameters,

A = model variance of t_i.

Level 1 accounts for the sampling variability of the direct estimates of true harvested yield, while level 2 links the true yield to known auxiliary variables (discussed in detail in Section 3). The smoothed variance estimates of the direct estimator are used to estimate the sampling variances $\{D_i\}$.

The Fay-Herriot model has been used extensively in small-area estimation and related applications due to its simplicity and ability to protect confidentiality of microdata and to produce design-consistent estimators. Some earlier applications of the Fay-Herriot model include estimation of: (i) false alarm probabilities in New York City (Carter and Rolph, 1974); (ii) batting averages of major league baseball players (Efron and Morris, 1975); and (iii) prevalence of toxoplasmosis in El Salvador (*Ibid.*). More recently, the Fay-Herriot model was used to estimate poverty rates for states, counties, and school districts in the US (Citro and Kalton, 2000) and to estimate proportions at the lowest level of literacy for states and counties (Mohadjer et al., 2007). In each case, the survey sample sizes in the areas of interest were insufficient to obtain direct estimates of adequate precision. A wide variety of methods has been developed to address such problems related to small-area estimation; see (for example) Lahiri (2003), Rao (2003, chapter 7), and Jiang and Lahiri (2006).

The Bayes estimator of t_i under squared error loss function is given by:

$$\hat{t}_i^{B} \equiv (1 - B_i)\hat{t}_i^{Dir} + B_i x_i^T \beta$$

where:

$$B_i = \frac{D_i}{A + D_i} \quad (i = 1, \cdots, m).$$

We first assume that the model variance *A* is known but the regression coefficients $\{\beta_i\}$ are unknown. The weighted least squares estimator of the vector of regression coefficients is given by:

$$\hat{\beta}(A) = \left(\sum_{j=1}^{m} \frac{1}{A + D_j} x_j x_j^T \right)^{-1} \sum_{j=1}^{m} \frac{1}{A + D_j} x_j \hat{t}_j^{Dir} .$$

An empirical Bayes (EB) estimator of true harvested yield can then be obtained by inserting this formula into the above expression for $\hat{t}_i^B$:

$$\tilde{t}_i^{EB} \equiv \tilde{t}_i^{EB}(A) = (1 - B_i)\hat{t}_i^{Dir} + B_i x_i^T \hat{\beta}(A).$$

The estimator $\tilde{t}_i^{EB}$ is also the *best linear unbiased predictor* (BLUP) estimator of t_i under the following simple linear mixed model:

$$\hat{t}_i^{Dir} = t_i + e_i = x_i^T \beta + v_i + e_i$$

where the sampling errors $\{e_i\}$ and county specific random effects $\{v_i\}$ are assumed to be independent with:

$$e_i \sim N[0, D_i],$$
$$v_i \sim N[0, A].$$

Thus the EB estimator shrinks the direct estimator $\tilde{t}_i^{Dir}$ toward the regression synthetic estimator $x_i^T \hat{\beta}(A)$, with the degree of shrinkage determined by B_i . The higher the value of B_i , the greater is the strength of the level 2 model and hence the efficiency of the EB estimator as reflected by a smaller value of its mean squared error (MSE). When $A = 0$, the level 2 model is perfect - $B_i = 1$ ($i=1,\ldots,m$). In this case, the EB estimator is identical to the regression synthetic estimator. However, this situation is unrealistic since level 2 modeling (like any modeling) cannot be perfect (i.e., *A* will always be strictly greater than *0*).

In practice, the model variance *A* (and hence the county specific shrinkage factors $\{B_i\}$) are unknown and must be estimated from the data. A number of consistent estimators of *A* have been proposed in the literature (see Rao, 2003). In the current application, we used the residual maximum likelihood (REML) method to estimate *A*. When the REML estimate of B_i is substituted into the above expression for $\tilde{t}_i^{EB}$, one obtains the following empirical Bayes estimator of t_i:

$$\hat{t}_i^{EB} \equiv \hat{t}_i^{EB}(\hat{A}) = (1 - \hat{B}_i)\hat{t}_i^{Dir} + \hat{B}_i x_i^T \hat{\beta}$$

where:

$\hat{\beta} = \tilde{\beta}(\hat{A})$,

$\hat{B}_i$ = REML estimate of B_i.

This estimator is also known as an *empirical best linear unbiased predictor* (EBLUP) of t_i.

The mean square error (MSE) of $\hat{t}_i^{EB}$ is defined as:

$$MSE(\hat{t}_i^{EB}) = E(\hat{t}_i^{EB} - t_i)^2$$

where E denotes the expectation with respect to the Fay-Herriot model. We propose to estimate this mean square error by the following parametric bootstrap estimator:

$$mse(\hat{t}_i^{EB}) = E_*(\hat{t}_i^{*EB} - t_i^*)^2$$

where $\hat{t}_i^{*EB}$ is exactly like $\hat{t}_i^{EB}$ except that it is based on the parametric bootstrap direct estimates $\hat{t}_i^{*Dir}$, and E_* is the expectation with respect to the following parametric bootstrap model that mimics the original Fay-Herriot model:

<u>Level 1</u> - $\hat{t}_i^{*Dir} \mid t_i^* \sim ind\, N[t_i^*, D_i]$,

<u>Level 2</u> - $t_i^* \sim ind\, N[x_i^T \hat{\beta}, \hat{A}]$ $\quad (i=1,\ldots,m)$.

A double bootstrap method along the lines of Chatterjee and Lahiri (2007) may be developed to correct for the potential bias in $mse(\hat{t}_i^{EB})$. But we believe that the extent of bias is likely to be small, especially if D_i/A is large. Further details on the parametric bootstrap MSE estimator can be found in an unpublished manuscript by Lahiri (2010).

3. Auxiliary Variables

In this section, we discuss the construction of auxiliary variables from different data sources. The objective is to improve the predictive power of the model by identifying a set of variables considered good potential predictors of harvested yield.

We first consider an auxiliary variable that uses historical data on official and previous census production statistics. Let $P_{it}^{(Off)}$ denote the official NASS estimate of crop production for county i in year t (t = 0, …, T), where $t=0$ represents the most recent census year and $t=T$ the current year. Let $P_{i0}^{(Cen)}$ denote the production for county i from the previous census. If this census figure is not available, it can often be imputed by multiplying the corresponding official production figure by the ratio $R_{(dist)}$ between the sums of census and official estimates for counties in the district where both figures are available. If both the census and official production are missing for a county, an imputed census value may still be generated (assuming all required numbers are available) by: 1) summing the non-missing official county level production estimates in the district, 2) subtracting that total from the official district level estimate (thereby obtaining a combined estimate for the counties with missing official production, 3) multiplying that number by $R_{(dist)}$ (to 'ratio up' to the census), and 4) multiplying the result by the ratio between number of (area frame) population units in the county of interest and total number of population units for all counties in the district with missing official production.

Let $\tilde{P}_{it}$ denote a new estimator of production for county *i* in year *t*. Then the following recursive relationship can be used to move the estimator forward:

$$\tilde{P}_{it} = [P_{it}^{(Off)} / P_{i,t-1}^{(Off)}]\tilde{P}_{i,t-1}, \ t=1,...,T-1$$

with $\tilde{P}_{i0} = P_{i0}^{(Cen)}$.

If either $P_{it}^{(Off)}$ or $P_{i,t-1}^{(Off)}$ is missing for a county, the ratio between official district level production estimates for years *t* and *t-1* can be substituted for $P_{it}^{(Off)} / P_{i,t-1}^{(Off)}$ in the above formula. In the event that the district level estimate is unavailable for at least one of the two years, the state level ratio between official production estimates could be used (although that was not necessary for the seven-state study discussed in the next section).

We shall adjust the estimates $\tilde{P}_{it}$ so that they sum to the official level state estimate of production. The resulting benchmarked estimates are given by:

$$\hat{P}_{it} = \tilde{P}_{it}[P_{state,t} / \sum_{j=1}^{m} \tilde{P}_{jt}],$$

where:

$P_{\text{state,t}}$ = official state level estimate of production for year *t*,

m = number of counties in the state.

Multiyear estimates of harvested acreage can be computed in similar manner, with multiyear estimates of harvested yield then obtained as the ratio between estimated production and estimated harvested acreage. In this paper, we consider the multiyear harvested yield for the previous year (2006 in the study) as an auxiliary variable in the Fay-Herriot model.

Nomalized Difference Vegetation Index (NDVI) figures are derived from Moderate Resolution Imaging Spectro-Radiometer (MODIS) data. MODIS is a payload scientific instrument launched into Earth orbit by NASA in 1999 on board the Terra (EOS AM) satellite and in 2002 on board the Aqua (EOS PM) satellite. Designed to measure large scale global dynamics such as changes in cloud cover and processes taking place in the oceans, on land and in the lower atmosphere, MODIS captures data in 36 spectral bands and varying spatial resolutions, imaging the entire Earth every one to two days and providing global coverage with a 15 acre ground sample resolution and 8 and 16 day temporal windows.

The NDVI is calculated from MODIS measurements of surface reflectance of visible and near infrared spectra. This index describes the vegetation condition of a crop from emergence to senescence, with the maximum value being nearly one for optimal vegetation cover. The NDVI is defined as follows:

NDVI = (Near Infrared – Visible) / (Near Infrared + Visible).

County level NDVI figures for the seven states of interest were obtained at 16 day intervals throughout 2007, but exploratory data analysis determined that the period between March 22 and August 13 (corresponding to the 7th through 16th satellite data capture dates) was adequate for purposes of this study. Two variables computed from the NDVI data were suggested by researchers at NASS who work with MODIS data as potentially good predictors of crop yield – the peak and weighted average NDVI defined as follows:

$$\rho(i) = \max(NDVI_{i,7}, NDV_{i,8}, \ldots, NDVI_{i,16}),$$

$$\omega(i) = [NDVI_{i,7} / 18] + [(NDVI_{i,8} + \ldots + NDVI_{i,15}) / 9] + [NDVI_{i,16} / 18] \qquad (i=1,\ldots,m)$$

where:

$NDV_{i,k}$ = NDVI value for county *i*, data capture date *k*.

The weighted average is proportional to a composite trapezoidal rule approximation of the area under a fitted curve of the NDVI values over the time window.

4. Results

The study area for evaluating the empirical Bayes estimator of harvested yield includes the top five soybean producing states in the US for the year 2007 – Iowa, Illinois, Minnesota, Indiana and Ohio - as well as Missouri (7th) and Kansas (11th). The SAS MIXED procedure was used to fit the EB model. The three independent variables used (described in Section 3) were - 1) multi-year estimate of harvested yield for 2006, 2) peak NDVI value (over the March 22 to August 13, 2007 period), and 3) weighted average NDVI over the same period. The dependent variable was the direct estimator of harvested yield for 2007. The estimated variance of the direct estimator was used to compute the weights. Counties with missing values for any one of the three independent variables were not used.

For each state in the study, the following five estimate types were compared – empirical Bayes (EB), Kott- Busselberg (KB), direct (DIR), Stasny-Goel (SG) and official NASS estimates (OFF). Five accuracy metrics were computed for each of the five types - average absolute deviation (AAD), average squared deviation (ASD), average absolute relative deviation (AARD), average squared relative deviation (ASRD) and percentage below census (PBC). The census figures for harvested yield in 2007 were regarded as 'truth' and used as the basis for measuring accuracy. AAD is the mean of absolute deviations between county estimates and corresponding 2007 census values, ASD the mean of squared deviations between estimates and census values, AARD the mean of ratios between absolute deviations and census values and ASRD the mean of squared ratios between absolute deviations and census values. PBC is the proportion of counties with estimate less than the corresponding 2007 census value. Values of PBC below (above) 0.5 suggest overestimation (underestimation) tendencies for an estimator. Table 1 shows the computed accuracy measures for all seven states.

The five estimate types were ranked from best or worst based on each of the five metrics (with the ranks for PBC computed based on absolute deviation from 0.5). Table 2 shows the average ranks (over the five metrics) by state and Table 3 the average ranks (over the seven states) by metric. Examination of Table 2 shows that EB had lowest (best) average rank (including the official figures) in four of the seven states (Illinois, Indiana, Iowa and Ohio), while SG had lowest average rank in Kansas and Missouri and lower average rank than EB, KB and DIR in Minnesota. In terms of the individual performance measures, Table 3 shows that EB had the best overall average rank for AAD, ASD and AARD, with SG and the direct estimator having the best average rank for ASRD and PBC, respectively.

Figures 1 and 2 are bar charts displaying the AARD and PBC (from Table 1) by state. Note the large AARD values for all four estimators (as well as official figures) in Kansas compared with the other six states in Figure 1. The tendencies of DIR, EB and SG to overestimate yield and KB to underestimate it are clearly discernible from Figure 2.

Figure 3 is a set of seven box plots (one for each state) showing relative deviations from 2007 census figures of the five estimate types being evaluated. Once again, Kansas stands out from the other states with its larger ranges for EB, KB, DIR and OFF.

5. Conclusion

The results discussed in Section 4 suggest that we have found reasonably powerful auxiliary data to explain soybean harvested yield at the county level In other words, we believe that the level 2 modeling is reasonable. However, outlier problems were observed in the direct estimates which impact the empirical Bayes based methodology considered in the paper. To further improve the methodology, the direct estimates may need to be treated for outliers. Alternatively, we may consider unit level modeling which does not make use of direct estimates.

Table 1. Estimation Accuracy for Harvested Yield

State	Estimator	Metric				
		AAD	ASD	AARD	ASRD	PBC
Illinois	EB	1.33	2.67	0.036	0.002	0.54
	KB	2.7	12.6	0.07	0.009	0.85
	Direct	1.41	3.1	0.038	0.003	0.51
	SG	1.35	2.72	0.034	0.002	0.39
	Official	1.82	5.18	0.048	0.004	0.42
Indiana	EB	1.31	2.94	0.034	0.002	0.37
	KB	3.02	16.6	0.075	0.01	0.82
	Direct	1.34	3.1	0.034	0.002	0.36
	SG	1.47	3.39	0.037	0.002	0.24
	Official	1.83	3.45	0.044	0.002	0.00
Iowa	EB	1.45	2.96	0.029	0.001	0.15
	KB	2.7	13.5	0.055	0.006	0.82
	Direct	1.55	3.5	0.031	0.001	0.2
	SG	1.93	4.87	0.039	0.002	0.11
	Official	2.12	5.94	0.043	0.002	0.08
Kansas	EB	4.25	37.5	0.128	0.032	0.35
	KB	5.13	63.0	0.155	0.053	0.49
	Direct	4.57	45.7	0.137	0.038	0.35
	SG	3.52	20.2	0.107	0.018	0.23
	Official	3.57	23.8	0.108	0.021	0.35
Minnesota	EB	1.61	9.38	0.046	0.008	0.45
	KB	3.46	26.0	0.095	0.022	0.85
	Direct	1.68	10.4	0.048	0.009	0.48
	SG	1.37	3.71	0.037	0.003	0.35
	Official	1.32	2.67	0.034	0.002	0.19
Missouri	EB	1.91	7.24	0.063	0.012	0.25
	KB	2.14	12.2	0.065	0.011	0.56
	Direct	2.01	8.07	0.061	0.008	0.27
	SG	1.94	6.58	0.06	0.007	0.24
	Official	2.02	7.43	0.064	0.009	0.13
Ohio	EB	2.08	7.79	0.054	0.009	0.23
	KB	3.82	24.5	0.091	0.015	0.73
	Direct	2.22	9.48	0.056	0.008	0.23
	SG	2.82	12.1	0.069	0.009	0.07
	Official	2.5	11.8	0.062	0.009	0.22

Table 2. Average Estimator Ranks by State for Harvested Yield

State	Estimator				
	EB	KB	Direct	SG	Official
Illinois	1.6	5	2.6	2	3.8
Indiana	1.2	4.8	2.2	3.2	3.6
Iowa	1.4	4.4	1.8	3.2	4.2
Kansas	2.9	4.2	3.7	1.8	2.4
Minnesota	2.8	5	3.4	2.2	1.6
Missouri	2.8	4	2.6	1.8	3.8
Ohio	1.5	4.2	1.9	4.2	3.2

Table 3. Average Estimator Ranks by Metric for Harvested Yield

Metric	Estimator				
	EB	KB	Direct	SG	Official
AAD	1.6	5	2.9	2.4	3.1
ASD	1.7	5	3	2.3	3
AARD	2	5	2.7	2.1	3.1
ASRD	2.6	4.9	2.7	2.3	2.6
PBC	2.3	2.7	1.7	4	4.3
All	2.0	4.5	2.6	2.6	3.2

Figure 1. Average Absolute Relative Deviation (AARD) of Estimators by State

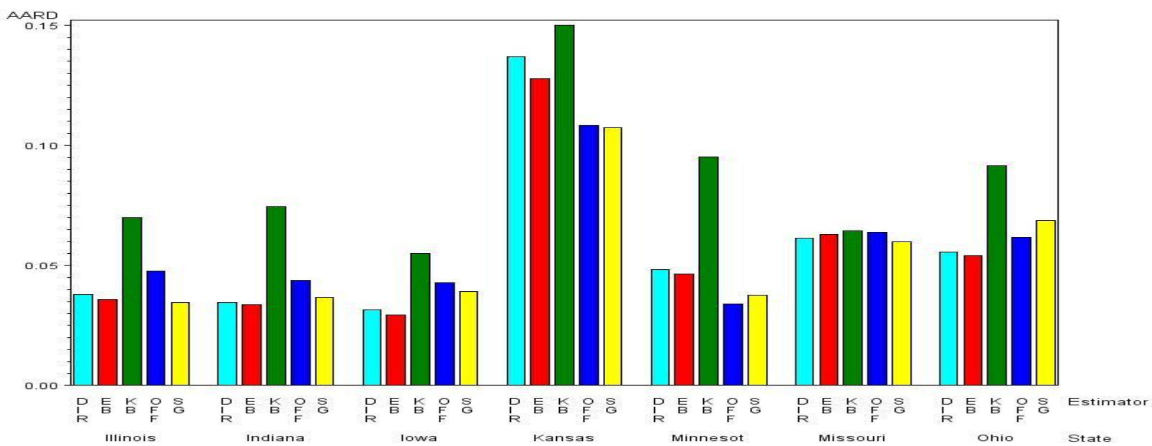


Figure 2. Percentage Below Census (PBC) of Estimators by State

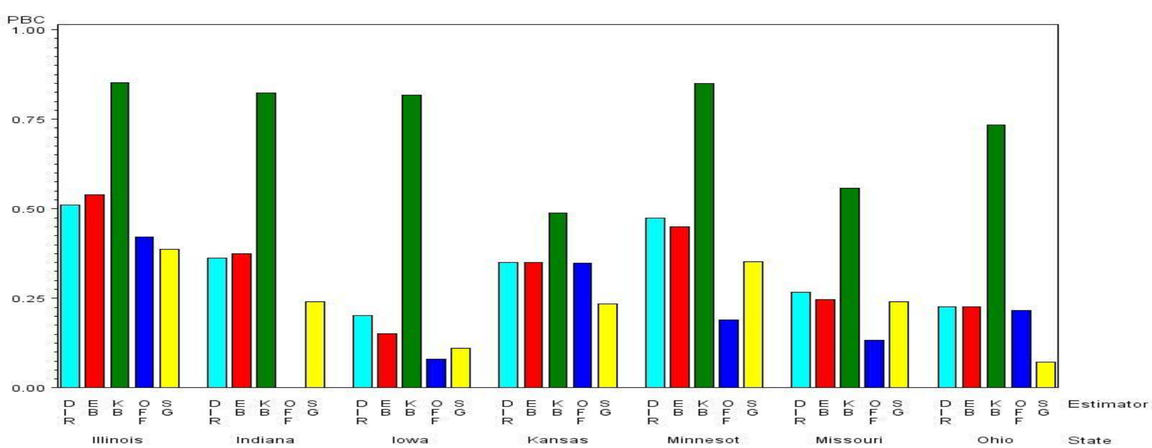
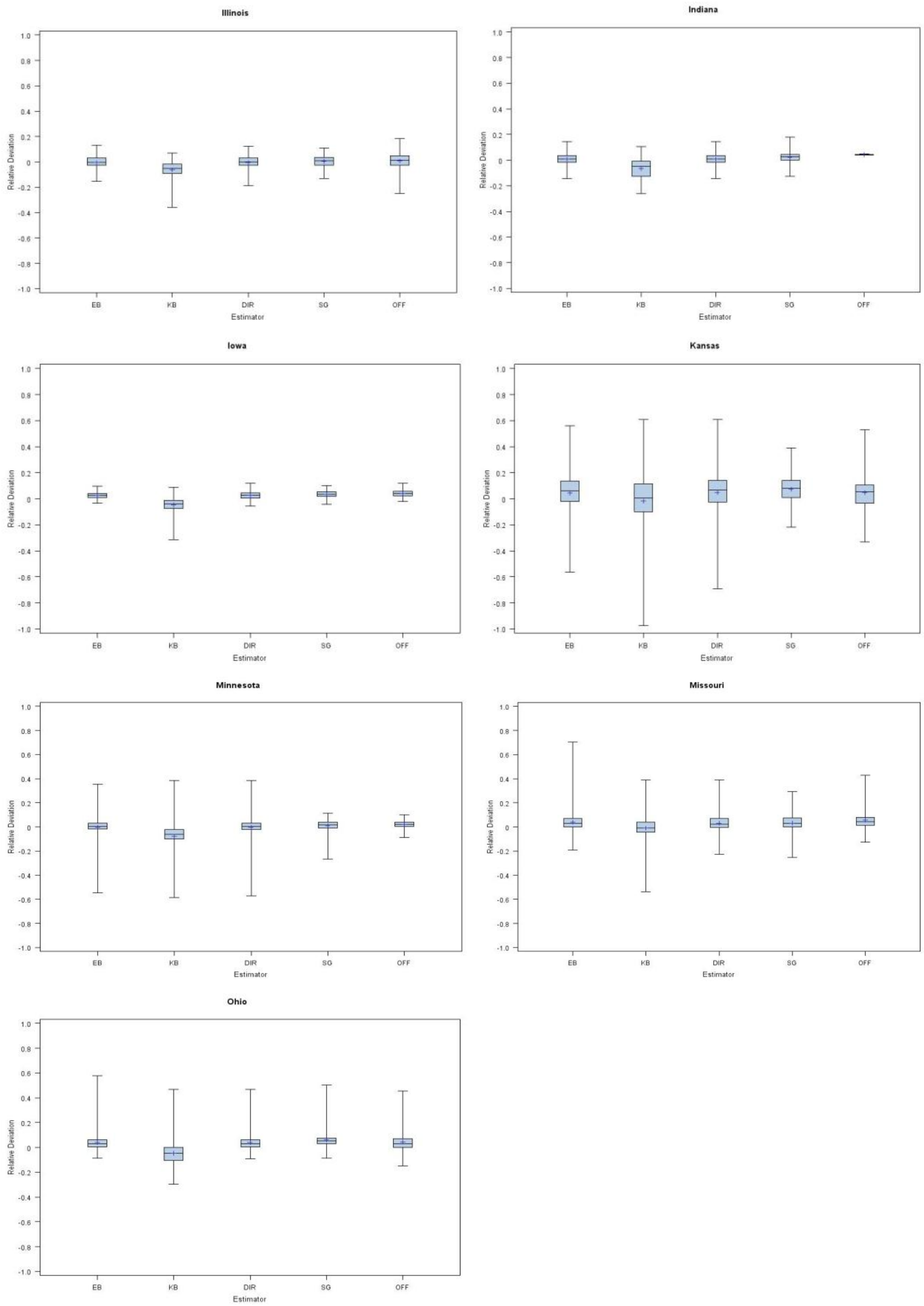


Figure 3. Box Plots of Relative Deviation of Estimates from Census 2007 Harvested Yield Figures



6. References

Bellow, M.E. (2007), *Comparison of Methods for Estimating Crop Yield at the County Level*. Research Report No. RDD-07-05, National Agricultural Statistics Service, U.S. Department of Agriculture.

Carter, G. and Rolph, J. (1974), Empirical Bayes Methods Applied to Estimating Fire Alarm Probabilities. *Journal of the American Statistical Association*, 69, 880-885.

Chatterjee, S. and Lahiri, P. (2007), “A Simple Computational Method for Estimating Mean Squared Prediction Error in General Small-Area Model”, Proceedings of the Section on Survey Research Methods, American Statistical Association, 3486-3493.

Citro, C., and Kalton, G. (eds.), (2000), “Small-Area Income and Poverty Estimates: Priorities for 2000 and Beyond”. Washington, DC: National Academy Press.

Crouse, C. (2000), *Evaluation of the Use of Spatial Modeling to Improve County Yield Estimation*. Research Report No. RD-00-05, National Agricultural Statistics Service, U.S. Department of Agriculture.

Efron, B., and Morris, C. (1975), Data Analysis Using Stein's Estimator and its Generalizations. *Journal of the American Statistical Association* 70, 311-9.

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

Griffith, D.A. (1999), *A Methodology for Small Area Estimation with Special Reference to a One-Number Agricultural Census and Confidentiality: Results for Selected Major Crops and States*. Research Report No. RD-99-04, National Agricultural Statistics Service, U.S. Department of Agriculture.

Griffith, D.A. (2001), *Progress Report, Part II: Model-Based Small Geographic Area Estimation: A Comparison of Alternative Methodologies*. Syracuse University Technical Report.

Iwig, W.C. (1993), “The National Agricultural Statistics Service County Estimates Program”, in *Indirect Estimators in Federal Programs*, Statistical Policy Working Paper 21, Subcommittee on Small Area Estimation, Federal Committee on Statistical Methodology, Office of Management and Budget.

Jiang, J., and Lahiri, P. (2006), Mixed Model Prediction and Small Area Estimation (with discussions), *Test. An Official Journal of the Spanish Society of Statistics and Operations Research*, 15, 1, 1-96.

Kott, P.S. (2009). “Some Ideas for a New Set of County-Estimates Crop Indications: an Update”, unpublished manuscript.

Lahiri, P. (2003), A Review of Empirical Best Linear Unbiased Prediction for the Fay-Herriot Small-Area Model, *The Philippine Statistician*, 52, nos. 1-4, 1-15.

Lahiri, P. (2010), “Some Thoughts on the Estimation of Crop Indications at the County Level”, unpublished manuscript.

Mohadjer, L., Rao, J.N.K., Liu, B., Krenzke, T., and Van de Kerckhove, W. (2007), Hierarchical Bayes Small Area Estimates of Adult Literacy Using Unmatched Sampling and Linking Models. Proceedings of the American Statistical Association.

Rao, J. N. K., (2003), *Small Area Estimation*, Wiley, New York.

Stasny, E. A., Goel, P.K., and Rumsey, D.J. (1991), County Estimates of Wheat Production, *Survey Methodology*, 17, 211-225.

Stasny, E. A., Goel, P.K., Cooley, C.A, and Bohn, L.L. (1995), "Modeling County-Level Crop Yield with Spatial Correlations Among Neighboring Counties". Department of Statistics. The Ohio State University.