Estimation of crop yield distribution and Insurance Premium using Shrinkage Estimator: A Hierarchical Bayes and Small Area Estimation Approach

Sebastain N. Awondo

Graduate Student
Department of Agricultural & Applied Economics, University of Georgia,
309 Conner Hall, Athens, Georgia, 30602
Email: sawondo@uga.edu

Gauri S. Datta

Professor
Department of Statistics, University of Georgia
226 Statistics Building, Athens, Georgia, 30602
Email: gauri@uga.edu

Octavio A. Ramirez

Professor and Head
Department of Agricultural & Applied Economics, University of Georgia
301 Conner Hall, Athens, Georgia, 30602
Email: oramirez@uga.edu

Esendugue Greg Fonsah

Associate Professor
Department of Agricultural & Applied Economics, University of Georgia
15 RDC Rd, Room 118, P.O. Box 1209, Tifton, Georgia, 31793
Email: gfonsah@uga.edu

Abstract

☆Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's Annual Meeting,Seattle, Washington, August 12-14, 2012.

Obtaining reliable estimates of insurance premiums is a critical step in risk sharing and risk transfer necessary to ensure solvency and continuity in crop insurance programs. Challenges encountered in the estimation include dealing with aggregation bias from using county level yield averages as well as properly accounting for spatial and temporal heterogeneity. In this study, we associate some of these challenges as classical small area estimation (SAE) problems. We employ a hierarchical Bayes (HB) SAE to obtain design consistent expected county level yields and Group Risk Plan (GRP) premiums for corm farms in Illinois using quasi-simulated data.

Preliminary results show little bias ($< 10\%$) in estimated expected county yields in several counties investigated. We found wide variation in GRP, APH and basis risk across counties for similar level of coverage and scale. Results show that farmers could lower their GRP premiums by as much as 30% by carefully choosing a coverage level and scale combination.

Keywords: Crop Insurance, Small area estimation, Hierarchical Bayes

1. Introduction

Crop Insurance is one of the most important tool for managing agricultural risk in the US and developed nations at large. Obtaining and using reliable estimates of insurance premiums is a critical step in risk sharing and risk transfer necessary to ensure solvency and continuity in crop insurance programs(Ramirez,1997; Ramirez et al, 2003; Ker and Coble 2003; Pope and Just 1999; Sherrick et al, 2004). Farm-level based and area-level based insurance policies are the two categories of policies offered by the Federal crop insurance program (FCIP) for selected crops in different regions of the country. Farm-level policies such as the traditional Actual Production History (APH) insure farmers against unpreventable losses and are designed for each farm based on their production history. Indemnities for the APH are triggered when the observed yield fall below the expected farm yield. On the other hand area-based policies introduced in the early nineties such as the Group Risk Plan (GRP) and the Group Risk Income Protection (GRIP) are designed to insure farmers against widespread or catastrophic losses. Indemnities for GRP are triggered when the observed county average yield falls below a trigger amount which is function of the expected county average yield, the coverage level and a scale. The expected farm and county yields are estimated by the national Agricultural statistical service (NASS) from historical farm yields and historical county-level averages respectively.

Major problems still exist in reliably estimating crop distributions and insurance premiums for both farm-level and area-level policies. Common problems encountered in their estimation include properly accounting for spatial and temporal effects (Kardner and Kramer, 1986; Ozaki et al, 2008;Claassen and Just,2009) as well as uncertainty in the parameter estimates. The absence of long historical yields on farms is the most common problem encountered in the estimation of farm-level policies potentially leading to highly biased estimates. Results from an empirically grounded simulation by Ramirez and Carpio (2011) showed that the high level of subsidy needed to keep the APH solvent can mostly be explained by the use of biased premium estimates and not adverse selection by farmers presumed to have a better knowledge about their risk exposure than the insurer. They also showed that actuarially fair premiums could be obtained by using improved estimation methods and large sample size.

On the other hand, estimation of area-level policies is generally plagued by aggregation bias resulting from the use of county-level averages. Aggregated county level data fails to explicitly represent systemic and random variation inherent with farm-level data. Claassen and Just (2009) found that using county-level averages in risk estimation understates farm-level yield variation by 50% and that 61% of systemic variation and 42% of random variation is lost by aggregating yields.This indicates strong need for alternative approaches to estimate risk premiums directly from farm-level data with minimum loss of information. Aggregation bias is likely to increase as yield data from fewer farms is used to construct county averages. Since the number of farms from which data is collected could vary from county to county and in some cases likely to be far smaller than the actual total number of farms within the county, the observed and hence expected county means derived using the county averages could be unreliable. This situation give rise to classic small area estimation (SAE) problems requiring the use of SAE methods to obtain more efficient parameter and premium estimates (Fay and Herriot, 1979; Datta and Ghosh, 1991; Datta et al,1999; Ghosh and Rao, 1994; Rao, 2003).

Small area estimation is an active area of research which involves obtaining reliable estimates from subpopulations (district, county, state, country, sex, race, sex-race combination, etc) when the survey data involves few observations at least in some subpopulation (commonly referred to as area). The methods developed circumvent this limitation by "borrowing strength" or making use of information from sample variables outside the area of interest. Typically sources from which strength is borrowed include data from neighboring or similar areas in which case we refer to as 'borrowing strength' across space and data from earlier time periods referred to as 'borrowing strength' across time. This process increases the 'effective sample' size use in the study (Datta and Ghosh, 1991;Rao, 2003) and thus the efficiency of estimated parameters.

Aggregation bias decreases the correlation between county-level and farm-level yield, thus increasing basis risk making area level insurance policies unreliable to farmers. This is because a farmer could incur significant yield losses and still not receives an indemnity. Likewise, a farmer could also be compensated for yield losses due to significant county-wide loss without actually experiencing significant yield loss on his farm. The difference between observed losses at the farm level and that at the county level is known as basis risk and constitute the main drawback in implementing and expanding area-based crop insurance programs despite the advantages it has over farm-level policies; area-based policies cost considerably less to administer with potential of reducing adverse selection and moral hazard. This is because claim agents are not required to carry out a damage assessment before issuing payments, and farmers are less likely not to know the true distribution of the county average yield thus preventing them from self-selecting into specific plans. In addition, incentives for farmers to engage in negligent behavior after obtaining coverage is significantly reduced since a poor yield on one or few farms may not be sufficient enough to lower the observed county average yield down to the trigger level. These advantages associated with area-based policies are translated into lower premiums compared to traditional farm level policy premiums.

In 2011, the risk management agency (RMA) responsible for administering the FCIP covered over 265 million acres, assuming over $80 million in liability. However, despite significantly lower GRP premiums, farmers overwhelmingly adhered to APH policy, with GRP contributing only 6% of the total FCIP liability (RMA, 2011). It is not clear how much of the adherence in APH is due to basis risk, inaccurate APH premiums and or individual risk behavior. To properly investigate the degree to which each of these factors contributes to the status-quo bias requires insight knowledge of the true distribution of the area and farm level premiums,basis risk as well as farmers' risk preferences. This study takes important steps towards achieving this goal by proposing alternative methods for estimating yield distributions and premium rates that minimizes bias from using aggregated data, accounts for uncertainty in parameter estimates while representing spatial and temporal heterogeneity.

We employ a two-step hierarchical Bayes (HB) estimator for SAE (Fay and Herriot, 1979;Datta and Ghosh, 1991; Datta et al,1999; Ghosh and Rao, 1994; Rao, 2003, You and Rao, 2003)to estimate yield distributions and actuarially fair premiums for GRP and APH policies for corn farms in 18 counties in Illinois. Hierarchical Bayes estimators directly accounts for uncertainty in parameter estimates and provide a convenient way of properly representing both spatial and temporal heterogeniety in the model. Moreover, significant progress in empirical Bayes and HB modeling together with advances in computational power facilitates obtaining stable HB estimators (Gelman and Rubin, 1992; Gelfand and Smith, 1990; Banerjee et al,2004).

Accounting for sample design in the estimation is necessary to obtain design consistent parameter and premium estimates. No risk analysis study that we know of has attempted to obtain design consistent estimates eventhough farm level data collected by NASS-USDA are likely to be design based or at least weighted to make sure large farms are more likely to be included in the sample. Our data generation and estimation accounts for sample design to ensure design consistent estimates. We circumvent dealing with limitation of obtaining farm level yields and covariates by harnessing the advantages offered by geospatial data and land observatory satellites (LANDSAT)to generate quasi-population and sample for corn yield. This approaches also offers the advantage that it allows us to compare estimates with true values and thus directly evaluate the efficiency gains of the model. Also our analysis could easily be extended to investigates the potential impact of catastrophic weather (in and out of sample) on yield losses and solvency of the FCIP based on model predictions as well as impact of alternative sampling approaches and estimation methods on efficiency of parameter estimates which we deal with in a separate paper.

The rest of the paper is organized as follows. We present our model use for estimation in section two. Section three handles the data generation process while section four presents results and discussions. Finally, we conclude with a summary of major findings.

2. Model specification and estimation

We propose a two-step hierarchical bayes estimator for small area means for unit level NER model. We begin by specifying the estimator for unit level NER model with cross sectional data following Prasad and Rao (1999) You and Rao(2003) and then proceed with extension for longitudinal data.

2.1. N.E.R model with cross-sectional data

The basic unit level NER model as specified by Batesse et al (1988) takes the form below.

$$y_{ij} = x_{ij}^T \beta + u_i + e_{ij}, j = 1, ..., n_i, i = 1, ..., m \tag{2.1}$$

Where y_{ij} is the response of unit j in area i, x_{ij} is the vector of auxiliary variables, β is the vector of fixed parameters, u_i is the random effect of area i and e_{ij} the random individual error term. The county effects u_i are assumed independent with zero mean and variance σ_u^2. Similarly, the errors e_{ij} are independent with mean zero and variance σ_e^2 , u_i's and the e_{ij}'s are assumed mutually independent.We can approximate the mean yield for county i at time t by (θ_i)[1].

$$\theta_i = \bar{X}_i^T \beta + u_i \tag{2.2}$$

Lets suppose that data was collected from n_i corn plots where each sample (n_i) is weighted by the area of the plots with weights $\breve{w}_{ij}$. We can combine equation 2.1 with the direct small area estimator ($\bar{y}_{iw}$) to produce an area-level NER model (2.3)[2].

$$\bar{y}_{iw} = \bar{x}_{iw}^T \beta + u_i + \bar{e}_{iw}, i = 1, ..., m \tag{2.3}$$

To develop an HB estimator based on equation 3.1, we consider that (i) $y_{ij}|\beta, u_i, \sigma_e^2 \sim N(x_{ij}^T \beta + u_i, \sigma_e^2), j = 1, ..., n_i,\ j = 1, ..., m$;(ii)$u_i|\sigma_u^2 \sim$ N(0,σ_u^2), and (iii) $\beta \sim$ N(0,H) where H is the variance covariance matrix of β. The precision parameter of each of the variance components is assumed to follow an inverse gamma distribution with different parameters; $\sigma_e^2 \sim IG(\lambda_1, \tau_1)$ and $\sigma_u^2 \sim IG(\lambda_2, \tau_2)$. The joint posterior distribution function is then given by 2.5

$$\begin{aligned} f(y_{ij}, j = 1, ..., n_i, i = 1, ..., m, \beta, \sigma_u^2, \sigma_e^2) = \\ \prod_{i=1}^{m} [\prod_{j=1}^{N_i} (\frac{1}{\sigma_e^2})^{\frac{1}{2}} e^{-\frac{1}{2\sigma_e^2}(y_{ij} - x_{ij}^T \beta - u_i)^2} (\frac{1}{\sigma_u^2})^{\frac{1}{2}} e^{-\frac{1}{2\sigma_u^2} u_i^2} \\ X \prod_{l=1}^{p} (\frac{1}{h_l^2})^{\frac{1}{2}} e^{-\frac{1}{2h_l^2}\beta_l^2} (\frac{1}{\sigma_e^2})^{\lambda_1 + 1} e^{-\frac{\tau_1}{\sigma_e^2}} (\frac{1}{\sigma_u^2})^{\lambda_2 + 1} e^{-\frac{\tau_2}{\sigma_u^2}} \end{aligned} \tag{2.4}$$

Solving for the marginal posterior distributions from 2.4 gives the following full conditionals.

$$\beta | y_{ij}, u_i, \sigma_e^2, \sigma_u^2 \sim N(\Lambda \sigma_e^2 \Sigma_{i=1}^m \Sigma_{j=1}^{n_i} (y_{ij} - u_i) x_{ij}, \Lambda) \tag{2.5}$$

Where $\Lambda = (\sigma_e^2 \Sigma_{i=1}^m \Sigma_{j=1}^{n_i} x_{ij} x_{ij}^T + H^{-1})^{-1}$.

$$u_i | y_{ij}, \beta, \sigma_e^2, \sigma_u^2 \sim N((n_i + \frac{\sigma_e^2}{\sigma_u^2})^{-1} \Sigma_{j=1}^{n_i} (y_{ij} - x_{ij}^T \beta), (\frac{n_i}{\sigma_e^2} + \frac{1}{\sigma_u^2})^{-1}) \tag{2.6}$$

---

[1]where $\bar{X}_i^T$ and x_{ij} are vectors both with dimensions kX1 and $\bar{X}_i^T = \Sigma_{j=1}^{N_i} \frac{x_{ij}}{N_i}$

[2]Where $\bar{y}_{iw} = \frac{\Sigma_{j=1}^{n_i} \breve{w}_{ij} y_{ij}}{\Sigma_{j=1}^{n_i} \breve{w}_{ij}} = \Sigma_{j=1}^{n_i} w_{ij} y_{ij}; w_{ij} = \frac{\breve{w}_{ij}}{\Sigma_{j=1}^{n_i} \breve{w}_{ij}} = \frac{\breve{w}_{ij}}{\breve{w}_i}$ and $\Sigma_{j=1}^{n_i} w_{ij} = 1$. Similarly $\bar{x}_{iw} = \Sigma_{j=1}^{n_i} w_{ij} x_{ij}$ $\bar{e}_{iw} = \Sigma_{j=1}^{n_i} w_{ij} e_{ij}$ with $E(\bar{e}_{iw}) = 0$ and Var$(\bar{e}_{iw}) = \sigma_e^2 T \Sigma_{j=1}^{n_i} w_{ij}^2 \equiv \varrho_i^2$

$$\sigma_e^2|y_{ij},\beta,u_i,\sigma_u^2 \sim IG(\lambda_1+\frac{1}{2}\Sigma_{i=1}^m n_i,\tau_1+\frac{1}{2}\Sigma_{i=1}^m\Sigma_{j=1}^{n_i}(y_{ij}-x_{ij}^T\beta-u_i)^2) \quad (2.7)$$

$$\sigma_u^2|y_{ij},\beta,u_i,\sigma_e^2 \sim IG(\lambda_2+\frac{m}{2},\tau_2+\frac{1}{2}\Sigma_{i=1}^m u_i^2) \quad (2.8)$$

However, we are interested in finding the expected county yield (θ_i) based on y_{iw}. Following the same HB framework using the area level model in equation 2.3 gives a similar conditional marginal posterior of $u_i|y_{ij},\beta,\sigma_e^2,\sigma_u^2 \sim N(q_{iw}\bar{y}_{iw}-\bar{x}_{iw}^T\beta),q_{iw}\varrho_i^2)$ where $q_{iw}=\frac{\sigma_u^2}{\sigma_u^2+\varrho_i^2}$ Combining the mean and variance of the posterior of u_i with equation 3.2 gives the conditional posterior mean of (θ_i) as follows.

$$E(\theta_i|\bar{y}_{iw},\beta,\sigma_e^2,\sigma_u^2)=q_{iw}\bar{y}_{iw}+(\bar{X}_i+q_{iw}\bar{x}_{iw})^T\beta \quad (2.9)$$

and variance $q_{iw}\varrho_i^2$ where β,σ_e^2 and σ_u^2 are drawn from the posterior distributions from the unit level model (2.1).

2.2. Estimation

We re-estimate the model in order to obtain consistent parameters. In the first stage of our estimation,equation 2.5 to 2.8 is use in Gibbs sampling (Gelfand and Smith, 1990) to simulate the marginal posterior distributions of β,u_i,σ_e^2 and σ_u^2. We assume non-informative priors on $\beta,\ v_i,\ \sigma_e^2,\ \sigma_v^2$ given as $\beta_p \sim N(0,10^4),p=1,...,13.,\ v_i \sim,\ \sigma_e^2 \sim IG(10^{-3},10^{-3}),i=1,...,m.,\ \sigma_v^2 \sim IG(10^{-3},10^{-3})$.

To estimate expected county yields we draw s samples, s=1,...,k of the parameters $(\beta^{(s)};\sigma_e^{2(s)};\sigma_v^{2(s)})$ from the simulated joint posterior distribution and use them in equation 2.9. Expected county yield is then obtained by averaging over the $\theta_i^{'}$s:

$$\hat{\theta}_i^{HB}=\frac{1}{s}\Sigma_{s=1}^k[q_{iw}\bar{y}_{iw}+(\bar{X}_i+q_{iw}\bar{x}_{iw})^T\beta] \quad (2.10)$$

Likewise, posterior variance of the expected county yield is obtained by drawing s samples from the joint posterior distributiona and using them in the variance formula ($q_{iw}\varrho_i^2$) and then taking the average. The indemnity for GRP for each county is then calculated as follows:

$$I_{iGRP}=max([\frac{\breve{\theta}_i-(\hat{\theta}_i^{HB})COV}{\hat{\theta}_i^{HB}COV}]\hat{\theta}_i^{HB}(scale),0) \quad (2.11)$$

Where $\breve{\theta}_i$ is the observed county average, COV and scale is the coverage level and scale chosen by the farmer. In this study we take COV from 70% to 95% in increment of 5% while scale is from 0.9 to 1.5 in increment of 0.1. We evaluate all(42) GRP policy coverage-scale combination in each county.

3. Data

Farm level data that allows for estimation of unit level models and conduct indepth analysis of this nature are rare to find. We circumvent this limitation by using quasi-simulated farm level yield data from 18 counties in Illinois. This data has the advantage that it is generated from true covariates attributed to specific corn farm plots from a known population. Moreover, the data generation and thus analysis accounts for sampling design which is important to obtain design consistent yield and premium estimates (Rao and You, 1999). No risk analysis study that we know of has attempted to obtain design consistent estimates eventhough farm level data collected by NASS-USDA are likely to be design based or at least weighted to make sure large farms are more likely to be included in the sample.

We use geospatial climate data from corn farms in Illinois with guided parameter estimates from previous studies to simulate empirically sound farm yields. First we use 2011 cropland data maps from NASS-USDA obtained from NASA LANDSAT to extract corn farm polygons within 18 counties in Illinois which make up Agricultural district 40 and 50. Note that the satellite uses a 250 meter resolution 16-day composite Normalized Difference Vegetation Index (NDVI) to classify crops with a statistical classification accuracy of up to 97% for heavily monocultivated areas like Illinois (NASS-USDA,2010). Figure 1 below illustrates classified corn farm polygons in a few neighboring counties within the districts. Using the coordinates of each plot, we obtained plot specific climate data from the PRISM website.

Data on each corn farm polygon include minimum and maximum monthly temperature and cumulative monthly precipitation from 1950 to 2011, elevation and area of polygon. To proceed we dropped all plots less than 40470 m^2 (10 acres). After creating weights for each plot by dividing each plot's area by the total area within the county it is located, we then carried out a weighted random sample of n_i corn farm plots by county where n_i is drawn from a uniform distribution with range 0 to 15. We simulated yields for corn plots using the regression model below.

$$y_{ij} = b_0 - .346P_5 + 10.463P_6 + 6.849P_7 - 0.523P_8 - 0.087P_5^2 - 0.903P_6^2 - 0.304P_7^2$$

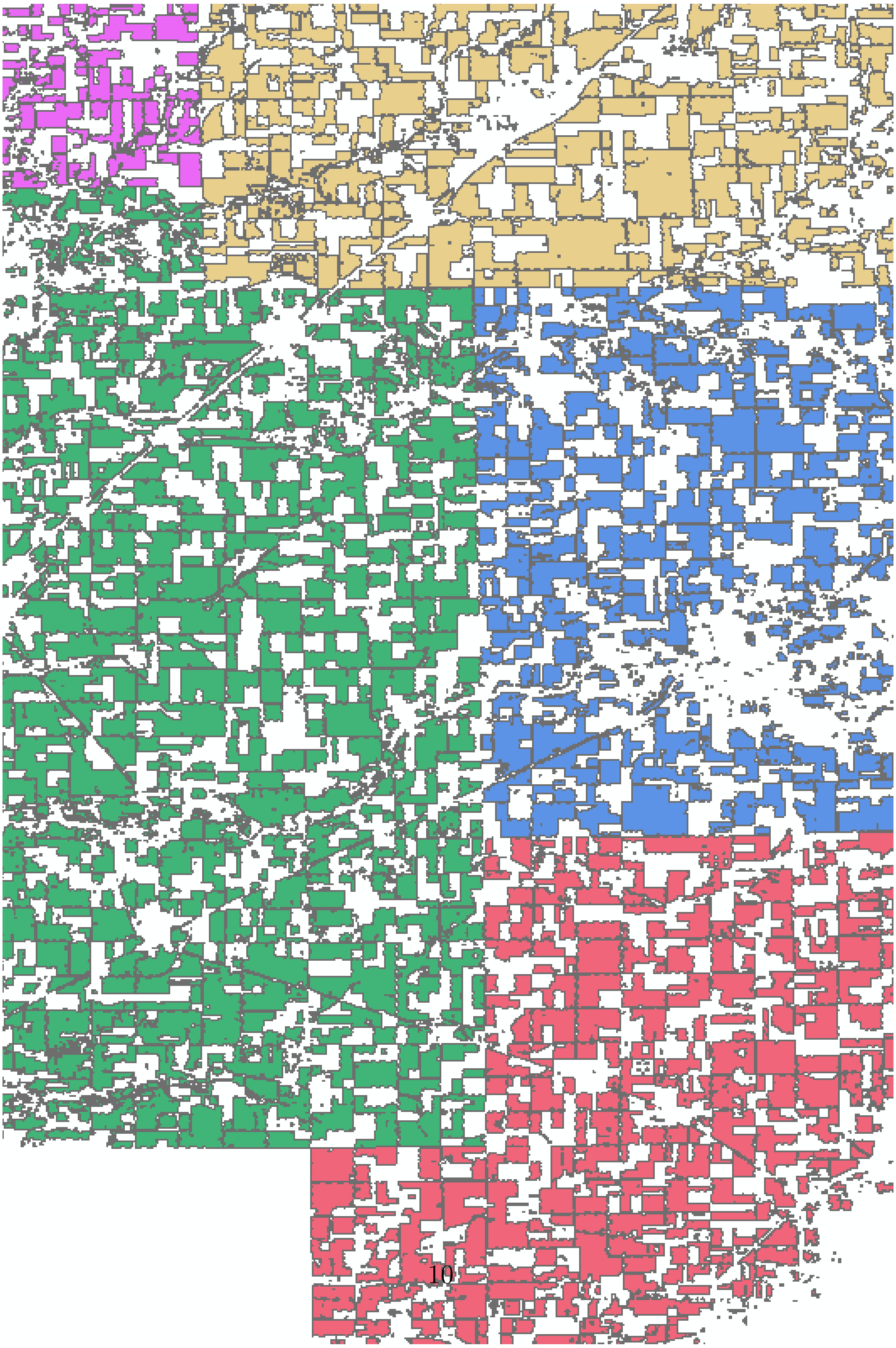

Figure 1: Corn plot polygons

$$+0.035P_8^2 + 1.232T_5 + 1.854T_6 - 2.013T_7 - 3.036T_8 + u_i + e_{ij} \quad (3.1)$$

Where y_{ij} is in bu/ha, P_5 to P_8 are cumulative precipitation (in) for May to August and P_5^2 to P_5^2 are their corresponding squares, T_5 to T_8 are temperatures (F) for May to August; b_0 is varied by county from 250 to 305; u_i is county random effect assumed to be normally distributed with mean 0 and variance 5 while e_{ij} is the error assumed to be normally distributed with mean 0 and variance 10. Our range of the two component variance is based on the range of aggregated variance estimated by Ramirez et al(2010) using farm level yields from endowment farms of the University of Illinois Urbana-Champaign. Also, our coefficient estimates are based on estimating the same model using detrended county level data. County level yields were obtained from NASS [3]. A similar regression model was used by Thompson (1988), Wolfram and Roberts (2006) and Tannura et al (2008) in investigating climate effect on yield variability.

3.1. Data summary

Table 1 and 2 shows summary of population and sample respectively based on data for 2011. 'A' represents area of plot in acres while Y represents yield per acre. Farm level yields range from 72 bu/ha in Mason to 222 bu/ha in Menard and Woodford. According to table 1, Stark county has the least total number of corn farm plots (507) while Livingstone has the most (2022). On the other hand summary of sample shows that as few as 1 corn farm is sampled in Livingstone while the largest samples are observed in Marshall, Logan and Kankakee. The sample shows variation in sample size across county with only few samples in some counties (such as 1 in Livingstone and 2 in Menard out of 2022 and 581 respectively). A direct county level average is likely to be unreliable especially in counties with very few observations. SAE methods borrows strength from similar areas with larger sample sizes to produce more reliable estimates. The sample yield range from 96 bu/ha in Mason to 206 bu/ha in Woodford. Sample county average yield is least for Mason (112) and highest for Menard (193) while the average plot size is least for Tazewell (299 acres) and highest for Marshall (547 acres)[4] . The range

[3]We thank Dr. Schlenker Wolfram for providing us with county level climate data

[4]The total number of corn plots within counties is different from the total number of corn farms from the same counties as given by 2007/2002 agricultural census. This is

Table 1: Data summary-population

County	Ni	mean(Y)	min(Y)	max(Y)	mean(A)	min(A)	max(A)
De Witt	706.00	164.31	131.94	200.62	159.93	10.01	1157.34
Logan	1023.00	169.31	132.21	204.97	208.07	10.01	1601.47
Macon	917.00	170.41	122.83	208.27	183.51	10.01	1233.40
Marshall	708.00	175.92	139.69	215.63	159.04	10.01	1188.26
Mason	871.00	111.98	72.35	148.92	151.91	10.01	2360.05
Mclean	1937.00	166.22	128.06	198.06	183.57	10.01	2040.70
Menard	581.00	179.79	149.37	222.20	141.78	10.01	2199.71
Peoria	1098.00	168.10	130.43	208.62	101.36	10.01	1095.96
Stark	507.00	169.25	126.39	203.13	194.47	10.01	1496.49
Tazewell	1245.00	165.80	128.00	200.63	131.43	10.01	1512.28
Woodford	999.00	186.10	153.38	222.19	143.10	10.01	1377.51
Champaign	1921.00	171.28	138.48	208.79	151.06	10.01	1149.11
Ford	873.00	124.27	78.39	156.27	172.85	10.01	1678.41
Iroquois	2041.00	135.44	87.46	179.53	171.02	10.01	2361.83
Kankakee	1112.00	146.02	110.28	187.58	181.51	10.01	1700.34
Livingstone	2022.00	168.97	125.36	210.89	155.76	10.01	1047.92
Piatt	733.00	163.35	124.91	192.67	191.18	10.01	1508.28
Vermillion	1608.00	122.29	84.77	158.82	140.34	10.01	1175.53

of the simulated yields and the differences in average yield across counties are comparable to the observed average yields published by NASS in the respective Illinois counties.

4. Results

These results are based on the unit level NER model using data for 2011 [5]. The summary of posterior distribution estimated by Gibbs sampling is shown in table 7. Table 3 compares estimated expected county level yields with their 'true' values. The bias($\hat{\theta}^{HB} - \theta$) from the estimated expected county level yield for all the counties except for Kankakee range between 1.7% in Mclean to 41% in Stark and Woodford. Seven counties (39%) have

partly due to that a farm could be made up of 2 or more corn plots

[5]Inferences from this model are still valid given that uncertainty in parameter estimates have been appropriately accounted for through the HB model

Table 2: Data summary-weighted sample

County	ni	mean(Y)	min(Y)	max(Y)	mean(A)	min(A)	max(A)
De Witt	12.00	165.27	151.26	192.37	421.11	64.27	773.47
Logan	14.00	168.60	148.75	195.13	339.57	14.68	928.94
Macon	12.00	170.13	152.17	197.37	391.00	29.58	677.64
Marshall	14.00	174.69	163.92	199.56	547.55	42.48	1188.26
Mason	9.00	112.04	96.78	126.36	404.20	86.51	999.00
Mclean	7.00	166.97	155.68	178.52	321.17	64.49	713.67
Menard	2.00	193.53	193.27	193.79	626.50	550.23	702.77
Peoria	11.00	163.97	149.78	177.40	400.11	65.38	1045.70
Stark	12.00	163.76	126.39	189.76	314.78	74.28	545.98
Tazewell	11.00	168.24	153.91	185.83	299.10	81.17	548.43
Woodford	12.00	188.64	173.19	206.15	351.14	15.79	982.99
Champaign	11.00	173.58	156.56	186.93	301.28	153.90	500.31
Ford	4.00	127.22	123.00	132.18	412.60	320.92	491.94
Iroquois	9.00	135.17	119.40	146.26	445.11	17.57	2361.83
Kankakee	14.00	148.27	129.51	173.42	366.93	45.59	1104.64
Livingstone	1.00	156.18	156.18	156.18	490.79	490.79	490.79
Piatt	9.00	164.52	146.85	180.42	445.63	12.01	1207.38
Vermillion	10.00	123.90	100.12	134.92	328.50	19.57	702.99

bias about 10% or less, four counties estimated with bias between 12% and 22% while 6 others have bias between 28% and 42%. Overall, the design based HB small area estimator appears to be efficient given the precision to which most of the expected county level yields have been estimated given the small sample sizes.

Livingstone and Menard which had sample sizes of 1 and 2 respectively have bias of 2% and 15% respectively. The unexpected high bias associated with the expected county yield for Kankakee could be due to outliers from parameter distribution.

Results for GRP indemnity under all possible coverage-scale scenarios are shown in table 4 for 6 of the counties. There appears to be considerable variation in indemnity (actuarially fair premium) across counties. Iroquois appear to have the lowest indemnities ($>$ 100%) compared to the other counties. GRP indemnity or actuarially fair premiums range from \$4/ha (for a 70% coverage and scale 0.9) to \$99/ha (for a 95% coverage and 1.5 scale) in

Table 3: Expected yield and Bias by county

County	θ	$\hat{\theta}^{HB}$	Stdv	Bias(%)
De Witt	164.31	158.76	0.0009	3.38
Logan	169.31	155.24	0.0005	8.31
Macon	170.41	141.67	0.0006	16.86
Marshall	175.92	244.68	0.0014	39.09
Mason	111.98	155.75	0.0007	39.08
Mclean	166.22	163.47	0.0002	1.65
Menard	179.79	151.08	0.0007	15.96
Peoria	168.10	146.63	0.0010	12.77
Stark	169.25	238.48	0.0008	40.91
Tazewell	165.80	159.51	0.0004	3.80
Woodford	186.10	109.19	0.0007	41.33
Champaign	171.28	156.91	0.0002	8.39
Ford	124.27	162.86	0.0004	31.05
Iroquois	135.44	165.10	0.0005	21.90
Kankakee	146.02	380.92	0.0005	160.86
Livingstone	168.97	165.52	0.0001	2.04
Piatt	163.35	146.98	0.0008	10.02
Vermillion	122.29	157.24	0.0004	28.59

Stark.

For the same level of coverage and scale, actuarially fair premiums for Marshall are \$0/ha and \$91/ha respectively. Premium rates for 95% coverage and 1.5 scale are considerably lower for Mason(\$57/ha), Ford(\$43/ha), Iroquois(\$34/ha) and Vermillion(\$40/ha). For the same coverage level, increasing the scale by a constant amount (e.g. 0.1) also increases the premium rate by a constant amount. However, increasing the coverage level while reducing the scale could either decrease or increase the premium, in some cases by considerable amounts. For example in Stark switching from a GRP policy with 70% coverage and 1.5 scale to a policy with 75% coverage and 0.9 scale results to an increase in premium by \$11.31/ha (166%). On the other hand switching from a GRP policy with a 90% coverage and 1.5 scale to a policy with 95% coverage and 0.9 scale results to a decrease in premium by \$25.29/ha (30%). Similar trends can be seen in Marshall county. These results shows that farmers can save significant amounts by careful selection of coverage and scale level.

5. Conclusion

Major challenges still exist is reliably estimating crop yield distributions and insurance premiums. These include properly accounting for uncertainty in parameter estimates, spatial and temporal heterogeneity, and dealing with aggregation bias stemming from using county level averages in estimation. In this study, we also view some of the problems as potentially stemming from the use of inadequate sample in generating county level yield averages and estimation of parameters. We employ a HB unit and area level SAE model to obtain more reliable expected county level yields and GRP premiums for 18 counties in Illinois using quasi-simulated data.

Preliminary results indicate little bias ($< 10\%$) in estimated expected county yields in several of the counties investigated. We found wide variation in GRP, APH and basis risk across counties for similar level of coverage and scale. GRP indemnities range from \$0/ha to \$100/ha. Farmers could lower their GRP premiums by as much as 30% and save a lot of money by carefully choosing a coverage level and scale combination.

6. References

Banerjee, S., B.P. Carlin and A.E. Gelfand. 2004. Hierachical Modeling and Analysis for Spatial Data. Chapman and Hall, CRC, London, Boca Raton, FL.

Battese, G.E., R.M. Harter and W.A. Fuller 1988.An error-components model for prediction of county crop areas using survey and satellite data.Journal of the American Statistical Association, 83, 2836.
Claassen, R., and R. E. Just. 2009. Heterogeneity and Distributional Forms of Farm-Level Yields. American Journal of Agricultural Economics 93(1): 144-160.
Datta, G.S. and M. Ghosh 1991. Bayesian prediction in linear models: applications to small area estimation. Annals of Statistics, 19, 17481770.
Datta, G.S. and P. Lahiri .2000. A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems.StatisticaSinica, 10, 613627.
Datta, G.S P. Lahiri and T. Maiti, and K.L. Lu. 1999. Hierarchical Bayes estimation of unemployment rates for the states of the U.S. Journal of the American Statistical Association, 94, 10741082.
Fay, R.E., and R.A. Herriot. 1979. Estimates of income for small places: an application of James-Stein procedures to census data. Journal of the American Statistical Association, 74, 269277.
Gardner, B. L., and R.A. Kramer. 1986. Experience with Crop Insurance Programs in the United States. In Crop Insurance for Agricultural Development: Issues and Experience, ed. P. Hazell, C. Pomareda, and A.Valdes,195222. Baltimore:Johns Hopkins University Press.
Gelfand, A.E., and A.F.M. Smith. 1990. Sampling-Based Approaches to Calculating Marginal Densities. Journal of the American Statistical Association 85:398409.
Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences (with discussion). Statist. Sci. 7, 457511.
Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: an appraisal. Statistical Science, 9, 5593.
Ghosh, M., N. Nangia, and D.H. Kim. 1996. Estimation of median income of four person families: a Bayesian time series approach. Journal of the American Statistical Association, 91, 14231431.
Ghosh, M., K. Natarajan, T.W.F. Stroud, and B.P. Carlin. 1998.Generalized linear models for small area estimation. Journal of the American Statistical Association, 93, 273282.
Just,R. E.,and R.D.Pope. 1999. Implications of Heterogeneity for Theory and Practice in Production Economics. American Journal of Agricultural Economics 81: 711718.
Ker, A.P., and K. Coble. 2003. Modeling Conditional Yield Densities. Amer-

ican Journal of Agricultural Economics 85:291304.
Ozaki, V. A., S. K. Ghosh, B. K. Goodwin and R. Shirota. 2008. Spatio-Temporal Modeling of Agricultural Yield Data with an Application to Pricng Crop Insurance Contracts. American Journal of Agricultural Economics 90(4) : 951961
Statist. Assoc. 85, 163171.
Prasad, N.G.N., J.N.K. Rao. 1999. On robust small area estimation using a simple random e.ects model. Surv. Methodol. 25, 6772.
Ramirez, O. A. 1997. Estimation and Use of a Multivariate Parametric Model for Simulating Heteroskedastic, Correlated, Nonnormal Random Variables: The Case of Corn Belt Corn, Soybean, and Wheat Yields. American Journal of Agricultural Economics 79: 191205.
Ramirez, O.A., S.K. Misra, and J. Nelson (2003). Efficient estimation of agricultural time series models with non-normal dependent variables. American Journal of Agricultural Economics 85(4) (November 2003): 1029-1040.
Ramirez, O.A. and C.A. Carpio. Premium estimation inaccuracy and the performance of the US crop insurance program Agricultural Finance Review (forthcoming in 2011).
Ramirez, O.A., T.U. McDonald and C.A. Carpio. A flexible parametric family for the modeling and simulation of yield distributions. Journal of Agricultural and Applied Economics 42,2(May 2010):1-17.
Rao, J.N.K. 1999. Some recent advances in model-based small area estimation Survey Methodology, 25, 175186.
Rao, J.N.K. 2003. Small Area Estimation. New York: Wiley.
Rao, J.N.K. and M. Yu. 1994. Small area estimation by combining time series and cross sectional data. Canadian Journal of Statistics, 22, 511528.
Risk Management Agency. 2011. Summary of Business Report and Data. U.S. Department of Agriculture (http://www.rma.usda.gov/data/sob.html)
Schlenker, W., and M. Roberts .2006.The Impact of Global Warming on U.S. Agriculture: An Econo-metric Analysis of Optimal Growing Conditions," Review of Economics and Statistics,88(1), 113-125.
Sherrick, B. J., F. C. Zanini, G. D. Schnitkey, and S. H. Irwin. 2004. Crop Insurance Valuation Under Alternative Yield Distributions. American Journal of Agricultural Economics 86: 406419.
Tannura, M. A., S. H. Irwin, and D. L. Good.Weather, Technology, and Corn and Soybean Yields in the U.S. Corn Belt. Marketing and Outlook Research Report 2008-01, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, February 2008. [$http$:

$//www.farmdoc.uiuc.edu/marketing/morr/morr_archive.html]$
Thompson, L.M. Effects of Changes in Climate and Weather Variability on the Yields of Corn and Soybeans. Journal of Production Agriculture 1(1988):20-27.
USDA-NASS. 2010. Cropland data layers
(http://www.nass.usda.gov/research/Cropland/Method/cropland.pdf)
You, Y. and J.N.K. Rao. 2003. Pseudo hierarchical Bayes small area estimation combining unit level models and survey weights. Journal of Statistical Planning and Inference 111 (2003) 197208.

Table 4: GRP indemnity under different Coverage-scale combination

Cov	scale	Marshall	Mason	Stark	Ford	Iroquois	Vermillion
0.70	0.90	0.00	0.00	4.08	0.00	0.00	0.00
0.70	1.00	0.00	0.00	4.53	0.00	0.00	0.00
0.70	1.10	0.00	0.00	4.98	0.00	0.00	0.00
0.70	1.20	0.00	0.00	5.44	0.00	0.00	0.00
0.70	1.30	0.00	0.00	5.89	0.00	0.00	0.00
0.70	1.40	0.00	0.00	6.34	0.00	0.00	0.00
0.70	1.50	0.00	0.00	6.80	0.00	0.00	0.00
0.75	0.90	10.58	5.73	18.11	0.00	0.00	0.00
0.75	1.00	11.76	6.36	20.13	0.00	0.00	0.00
0.75	1.10	12.94	7.00	22.14	0.00	0.00	0.00
0.75	1.20	14.11	7.64	24.15	0.00	0.00	0.00
0.75	1.30	15.29	8.27	26.16	0.00	0.00	0.00
0.75	1.40	16.46	8.91	28.18	0.00	0.00	0.00
0.75	1.50	17.64	9.54	30.19	0.00	0.00	0.00
0.80	0.90	23.69	14.13	30.40	3.45	0.00	2.13
0.80	1.00	26.32	15.70	33.77	3.83	0.00	2.37
0.80	1.10	28.95	17.27	37.15	4.21	0.00	2.61
0.80	1.20	31.58	18.84	40.53	4.59	0.00	2.85
0.80	1.30	34.21	20.41	43.91	4.98	0.00	3.08
0.80	1.40	36.84	21.98	47.28	5.36	0.00	3.32
0.80	1.50	39.48	23.55	50.66	5.74	0.00	3.56
0.85	0.90	35.25	21.54	41.23	11.86	5.47	10.33
0.85	1.00	39.16	23.94	45.82	13.18	6.07	11.48
0.85	1.10	43.08	26.33	50.40	14.50	6.68	12.63
0.85	1.20	46.99	28.73	54.98	15.82	7.29	13.78
0.85	1.30	50.91	31.12	59.56	17.14	7.90	14.93
0.85	1.40	54.83	33.51	64.14	18.46	8.50	16.07
0.85	1.50	58.74	35.91	68.72	19.77	9.11	17.22
0.90	0.90	45.52	28.13	50.87	19.35	13.42	17.62
0.90	1.00	50.58	31.26	56.52	21.50	14.91	19.58
0.90	1.10	55.64	34.39	62.17	23.65	16.40	21.54
0.90	1.20	60.70	37.51	67.82	25.80	17.89	23.50
0.90	1.30	65.75	40.64	73.47	27.95	19.38	25.45
0.90	1.40	70.81	43.76	79.13	30.10	20.87	27.41
0.90	1.50	75.87	46.89	84.78	32.25	22.36	29.37
0.95	0.90	54.72	34.03	59.49	26.04	20.53	24.14
0.95	1.00	60.80	37.81	66.10	28.94	22.81	26.83
0.95	1.10	66.87	41.59	72.71	31.83	25.09	29.51
0.95	1.20	72.95	45.38	79.32	34.73	27.38	32.19
0.95	1.30	79.03	49.16	85.92	37.62	29.66	34.87
0.95	1.40	85.11	52.94	92.53	40.51	31.94	37.56
0.95	1.50	91.19	56.72	99.14	43.41	34.22	40.24

Table 5: Summary posterior distribution

Parameter	mean	sd	MC error	2.5%	median	97.5%
b0	150.6	317.9	12.08	-519.7	178.9	734.6
P_5	388.9	246.3	8.246	-92.54	392.7	860.6
P_6	-144.0	24.55	0.7315	-192.4	-144.3	-95.47
P_7	-178.0	162.4	5.574	-479.7	-186.5	162.2
P_8	-162.9	68.15	2.139	-294.1	-162.3	-34.37
P_5^2	-55.41	27.4	0.8989	-107.8	-55.66	-0.5274
P_6^2	10.19	1.793	0.05349	6.721	10.23	13.77
P_7^2	54.98	48.31	1.693	-45.53	56.86	142.5
P_8^2	80.22	32.07	0.9625	20.26	80.03	142.5
T_5	-0.4338	17.02	0.54	-33.81	-0.2824	32.37
T_5	7.327	25.74	0.8464	-42.1	7.666	57.76
T_5	-24.06	25.35	0.9217	-72.19	-24.71	27.66
T_5	19.88	20.8	0.6524	-21.48	19.68	61.89
$\sigma^2_{u(1)}$	5.117	2.287	0.06613	1.68	4.85	10.71
$\sigma^2_{u(2)}$	5.027	2.304	0.06342	1.601	4.721	10.43
$\sigma^2_{u(3)}$	4.955	2.229	0.0779	1.558	4.597	10.37
$\sigma^2_{u(4)}$	4.883	2.137	0.07026	1.638	4.58	9.998
$\sigma^2_{u(5)}$	4.961	2.187	0.07663	1.575	4.707	9.883
$\sigma^2_{u(6)}$	4.897	2.233	0.0653	1.642	4.476	10.14
$\sigma^2_{u(7)}$	4.975	2.247	0.07123	1.582	4.625	10.23
$\sigma^2_{u(8)}$	4.86	2.043	0.06438	1.687	4.563	9.841
$\sigma^2_{u(9)}$	5.05	2.342	0.06577	1.622	4.613	10.43
$\sigma^2_{u(10)}$	4.906	2.12	0.06426	1.629	4.617	10.04
$\sigma^2_{u(11)}$	5.004	2.222	0.08111	1.759	4.713	10.55
$\sigma^2_{u(12)}$	4.993	2.282	0.06329	1.592	4.651	10.3
$\sigma^2_{u(13)}$	4.987	2.202	0.05626	1.613	4.658	9.894
$\sigma^2_{u(14)}$	5.026	2.177	0.06072	1.501	4.765	10.06
$\sigma^2_{u(15)}$	4.926	2.2	0.0786	1.479	4.542	9.867
$\sigma^2_{u(16)}$	5.118	2.302	0.0686	1.6	4.821	10.33
$\sigma^2_{u(17)}$	4.957	2.324	0.06199	1.687	4.503	10.78
$\sigma^2_{u(18)}$	5.098	2.232	0.07657	1.692	4.756	9.974
σ^2_e	0.004246	4.424E-4	1.381E-5	0.003412	0.004228	0.005163
deviance	1466.0	5.338	0.1714	1457.0	1465.0	1477.0

1=Champaign;2=De Witt;3=Ford;4=Iroquois;5=Kankakee;6=Livingstone;7=Logan; 8=Mclean;9=Macon;10=Marshall;11=Mason;12=Menard;13=Peoria;14=Piatt; 15=Stark;16=Tazewell;17=Vermillion;18=Woodford.