Small Area Estimation of Insurance Premiums and Basis Risk

Sebastain N. Awondo

Ph.D Candidate
Department of Agricultural & Applied Economics, University of Georgia,
309 Conner Hall; Athens, Georgia, 30602; USA; Email: sawondo@uga.edu

Gauri Datta

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

Octavio A. Ramirez

Professor and head
Department of Agricultural & Applied Economics, University of Georgia
301 Conner Hall; Athens, Georgia, 30605; USA; Email: oramirez@uga.edu

Esendugue Greg Fonsah

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

Abstract

The magnitude of basis risk between Actual Production History (APH) and Group Risk Plan (GRP) contracts across corn farms in Illinois counties is estimated using pseudo-simulated yields with farm specific geospatial climate data. A two-step hierarchical Bayes small area estimator was used to address problems related to lack of representative sample, aggregation bias, properly accounting for spatial and temporal heterogeneity and uncertainty in parameter estimates.

We found wide variation in expected basis risk across farms within and between counties. Expected basis risk was found to sharply increase under APH plans with higher coverage levels.

Keywords: Crop Insurance, Basis risk, Small area estimation, Hierarchical Bayes

☆Selected Paper prepared for presentation at the Southern Agricultural Economics Association Annual Meeting, Orlando, FL, February 4-7, 2013.

1. Introduction and Background

Area level crop insurance and farm level crop insurance are two major categories of crop insurance provided to farmers by the Federal Crop Insurance Program (FCIP). Indemnities for farm level policies are triggered when the observed yield falls below the yield guarantee.Area level policies are designed to insure farmers against widespread or catastrophic losses and indemnities are triggered when the observed county average yield falls below a trigger amount (yield guarantee). Area-based policies cost considerably less to administer with the 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 to know the true distribution of the expected county 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 to lower the observed county average yield down to the trigger level. The low cost of administration makes premium for area level policies considerably lower compared to premiums for farm level policies. However, a disproportionate amount of farmers prefer farm level policies such Actual Production History (APH) over area level policies such as Group Risk Plan (GRP) for farm risk management.

In 2011, the risk management agency (RMA) responsible for administering the FCIP covered over 265 million acres, assuming over $80 million in liability. However only 6% of the total FCIP liability (RMA, 2011) was attributed to GRP. Past studies have attributed the observed behavior to basis risk. Basis risk result from the lack of correlation between farm level crop losses and county level crop losses.

This study attempt to quantify the magnitude of basis risk involved across corn farms in Illinois counties using pseudo-simulated yields with farm specific geospatial climate data. To reliably estimate basis risk one must first efficiently rate contracts for the area and farm level policy involved.

Challenges faced by researchers to reliably estimate premiums still prevails and include making distributional form assumptions, properly accounting for spatial and temporal heterogeneity, lack of representative sample and dealing with aggregation bias (Just and Pope, 1999; Ozaki et al, 2008; Claassen and Just,2009). Recent findings from an empirically grounded simulation by Ramirez and Carpio (2011) showed that the high level of subsidy needed to keep the Federal Crop Insurance Program (FCIP) 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.

Following Awondo et al. (2012), we treat the lack of a representative sample as a classic small area estimation problem and employ a two-step hierarchical Bayes model (Fay and Herriot, 1979; Datta and Ghosh, 1991; Datta et al., 1999; Prasad and Rao, 1990; Ghosh and Rao, 1994; Prasad and Rao, 1999; You and Rao, 2003) that combines direct county estimates and county-level data to obtain more efficient expected county yields and expected loss (actuarially fair premium) for GRP policy. The resulting estimate is either shrunken toward or away from the direct estimate (simple county averages). Next we simulated actuarially fair premiums for Actual Production History (APH) and proceed to derive expected basis risk based on GRP and APH premiums.

Small area estimation is an active area of research aimed at obtaining reliable estimates from subpopulation (district, county, state, country, sex, race, sex-race combination, etc) when the data has few observations at least in some subpopulation (Datta and Ghosh, 1991; Datta et al., 1996, 2000, 2002; Rao, 1999, 2003). Suppose that the number of farms from which data is collected vary from county to county and in some cases likely to be far smaller than the actual total number of farms within the county. In this case, the observed county average yield and hence expected county yield based on the data could be unreliable giving rise to a classic SAE problem.

Following Awondo et al. (2012), we derived design consistent estimates under both simple random sampling (with equal weights on each observation) and weighted random sampling (with unequal weights) to accounts for the design. USDA-NASS area frame design for agricultural surveys which is the methodology used by NASS to develop and sample Primary Sampling Units (PSUs) and segments sometimes chooses segments based on the Probability of Selection Proportional to Size (PPS)(USDA-NASS, 2009). Thus putting more weight on large farms within a county. In this case, failure to account for the sampling design could lead to design inconsistent estimates (Datta et al., 1996, 2000; Prasad and Rao, 1999; You and Rao, 2003).

The rest of the paper is organized as follows. In section two, we specify the model use for estimation. Section three discusses the data and data generation process while section four presents results and discussions. Finally, we conclude with a summary of major findings and suggestions for future research.

2. Model specification and estimation

We specify a two-step hierarchical Bayes small area estimator for expected corn yields at county level. For simplicity, we use a nested error regression (NER) model with cross sectional data for both farm level and county level. The specification can easily be extended to longitudinal and time series data to fully represent temporal effects following Ghosh et al. (1996); Datta et al. (1999, 2002) and Torabi (2012). The model develop is based on the basic unit level NER model by Battese et al. (1988) and extensions by Prasad and Rao (1999) and You and Rao (2003).

The basic unit level NER model takes the form.

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

Where y_{ij} is the yield on farm j in county 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.If N_i is large,$N_i^{-1}\Sigma_{j=1}^{N_i} e_{ij} \approx 0$ and we can approximate the mean yield for county i by θ_i[1].

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

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

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

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

2.1. Hierarchical Bayes model

To develop an HB estimator based on equation 2.1, 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 equation 2.4.

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

Solving for the marginal posterior distributions from equation 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 $\Lambda = (\sigma_e^{-2}\Sigma_{i=1}^m\Sigma_{j=1}^{n_i}x_{ij}x_{ij}^T + H^{-1})^{-1}$.

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

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

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

Following the same HB framework using the area level model in equation 2.3 gives a similar conditional marginal posterior of $u_i|\bar{y}_{ij}, \beta, \sigma_e^2, \sigma_u^2 \sim N(q_{iw}(\bar{y}_{iw} - \bar{x}_{iw}^T\beta), q_{iw}\varrho_i^2)$ where $q_{iw} = \frac{\sigma_u^2}{\sigma_u^2 + \varrho_i^2}$ Combining the mean and variance of the conditional marginal posterior of u_i with equation 2.2 gives the conditional posterior mean of θ_i (equation 2.9) and variance $q_{iw}\varrho_i^2$.

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

where β, σ_e^2 and σ_u^2 are drawn from the joint posterior distributions derived from the unit level model (equation 2.1).

[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 $\text{Var}(\bar{e}_{iw}) = \sigma_e^2\Sigma_{j=1}^{n_i}w_{ij}^2 \equiv \varrho_i^2$

2.2. Estimation

In the first stage of our estimation, equation 2.5 to 2.8 is used 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 $\beta_p \sim N(0, 10^4), p = 1, ..., 13.,\ v_i \sim$ $,\ \sigma_e^2 \sim IG(10^{-3}, 10^{-3}), i = 1, ..., m.,\ \sigma_v^2 \sim IG(10^{-3}, 10^{-3})$. With initial values for $v_i, \sigma_e^2, \sigma_v^2$, we draw β from 2.5. Using the drawn β and initial values for σ_e^2, σ_v^2, we draw and update v_i with 2.6. Similarly, we draw and update σ_e^2 conditional on initial σ_u^2 and updated values of β, u_i. Finally, we also draw and update σ_u^2 given new values of β, u_i and σ_e^2 to complete single phase of simulation. The process is repeated 10000 times to produce 10000 draws for each conditional marginal posterior and the first 5000 draws were burnt. Three separate chains were simultaneously simulated each with 10000 draws and a burn-in of 5000. Diagnostic plots of the three chains are done to ensure convergence in the posterior distributions.

To estimate expected county yields we draw s samples of the parameters with replacement, s=1,...,k $(\beta^{(s)}; \sigma_e^{2(s)}; \sigma_v^{2(s)})$ 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:

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

Likewise, posterior variance of the expected county yield is obtained by drawing s samples from the joint posterior distribution and using them in the variance formula $(q_{iw}\varrho_i^2)$ and then taking the average. The same results can be obtained by simply finding the variance of the s simulated county mean draws for each county.

To derive expected loss (actuarially fair premium) in each county under the GRP plan, we use the direct county average yield from the sample together with s (1000) draws of θ_i to estimate losses for all coverage-scale combinations using equation 2.11. The expected yield loss ($bu/acre$) for each coverage-scale combination (L_{izp}^{GRP}) is taken as the average over the s simulations.

$$L_{izp}^{GRP} = \frac{1}{s}\Sigma_{s=1}^{k}[max([\frac{(\hat{\theta}_i^{HB})C_z - \bar{y}_{iw}}{\hat{\theta}_i^{HB}C_z}]\hat{\theta}_i^{HB}S_p, 0)], i = 1, ..., m\ z = 1, ..., 6\ p = 1, ..., 7 \tag{2.11}$$

Where $\bar{y}_{iw}$ is the direct county estimate. Note that in the case where the data is obtained by a simple random sampling $\bar{y}_{iw} = \bar{y}_i$. The coverage level (C_z) and scale (S_p) are chosen by the farmer. In this study we take C_z=(70%, 75%, 80%, 85%, 90%, 95%) and following Wang et al. (2012), we included a scale with values ranging from 0.9 to 1.0 in increment of 0.1 (i.e. S_p=(0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5)) to investigate its effect on expected loss under GRP and its potential to reduce basis risk. The reason for introducing the scale is to allow farmers to adjust county level expected loss to better correlate with the farm level expected loss. A scale of 0.9 will reduce the yield loss by about 10% while a scale of 1.2 will increase the yield loss by about 20%. A scale of 1 has no effect on the expected yield loss. We evaluate expected losses under the GRP for all(42) coverage-scale combinations in each county. Indemnity payments are normally derived by multiplying the percentage yield shortfall by a chosen proportion (0.6 - 1.0) of the maximum liability per acre.

The expected yield loss for APH is also simulated for each farm in the sample. In each case, farm level climate covariates are combined with 1000 draws of parameters from the joint posterior to produce 1000 farm level mean ($\hat{y}_{ij}$). The simulated means are then used with the direct farm yield (y_{ijw}) in equation 2.12 to produce estimated losses under the APH. The expected loss (actuarially fair premium) under APH in $bu/acre$ is then obtained by taking the average of the simulated losses as follows.

$$L_{ijq}^{APH} = \frac{1}{s}\Sigma_{s=1}^{k}[max([\hat{y}_{ij}C_q - y_{ij}^o], 0)], i = 1, ..., m, j = 1, ..., n_i, q = 1, ..., 6 \tag{2.12}$$

Where y_{ij}^o is the observed farm level yield and C_q=(70%, 75%, 80%, 85%, 90%, 95%).

Similarly, we estimate expected loss at the farm level not covered by the GRP policy (I_{ijpqz}^{BR}) on each farm for all coverage-scale combinations of the GRP and coverage levels of APH as follows.

$$L_{ijpqz}^{BR} = \frac{1}{s}\Sigma_{s=1}^{k}[max([I_{ijq}^{APH} - I_{izp}^{GRP}], 0)], i = 1, ..., m, j = 1, ..., n_i \tag{2.13}$$

Where p=1,...,7,q, z=1,...,6 and q and z can be the same or different. These estimates represent actuarially fair or net premiums for a potential insurance policy which can be introduced to cater for uninsured yield loss at farm level for GRP policy holders.

We estimated two models at all levels of our analysis. The first model is based on data generated from a weighted random sampling with unequal weights allocated to the farms sampled within each county while the second model is based on data generated from a simple random sampling with equal weights allocated to the farms sampled within each county.

3. Data

Following Awondo et al. (2012), we simulated corn farm yields using geospatial climate data specific to each corn farm and 'true' parameter estimates.

First, we used the 2011 crop classification map from USDA-NASS for the State of Illinois obtained with LANDSAT to identify and extract corn pixels in 18 counties which make up Agricultural district 40 and 50 [3]. The road data from the Illinois Department of Transportation was used to overlay these results and increase plot separability. The area of each plot was calculated and plots with area below two Landsat pixels ($1800m^2$) were dropped from the data set. Fields were added to the data, including a unique identifier for each plot. Climate data (images representing precipitation, maximum temperature and minimum temperature) from PRISM were imported, reprojected and resampled to match the projection, resolution and extent of the crop data. Monthly climate attributes for each plot were then extracted. These include minimum temperature, maximum temperature and cumulative precipitation. The average monthly temperature in each plot was calculated as the average of the minimum and maximum temperatures for the month.

All plots less than 10 acres (40470 m^2) were dropped. After creating weights for each plot by dividing each plot's area by the total area within the county it is located, we carried out a weighted random sample of n_i corn farm plots by county where n_i is drawn from a uniform

[3]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

distribution between 1 and 5. We simulated yields for corn plots using the regression model below.

$$y_{ij} = 320 - .346P_{ij5} + 10.463P_{ij6} + 6.849P_{ij7} - 0.523P_{ij8} - 0.087P_{ij5}^2 - 0.903P_{ij6}^2 - 0.304P_{ij7}^2$$

$$+0.035P_{ij8}^2 + 1.232T_{ij5} + 1.854T_{ij6} - 2.013T_{ij7} - 3.036T_{ij8} + u_i + e_{ij} \tag{3.1}$$

Where y_{ij} is in bu/acre, P_{ij5} to P_{ij8} are cumulative monthly precipitation (inches) for farm j in county i from May to August and P_{ij5}^2 to P_{ij5}^2 are their corresponding squares, T_{ij5} to T_{ij8} are average monthly temperatures (F) from May to August; u_i is county random effect assumed to be normally distributed with mean 0 and variance 15 while e_{ij} is the error assumed to be normally distributed with mean 0 and variance 25. Our range of variance components is consistent with the range 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. A similar regression model was used by Thompson (1988), Schlenker and Roberts (2006) and Tannura et al. (2008) and has been found to explain over 75% of the variability in corn yield.

3.1. Data summary

Table 1 shows a summary of the sample of corn farms in Illinois counties used in the estimation. The results show that Iroquois has the most number of corn farm population (2041) while Stark has the least amount (507). The number of farms sampled in each county

Table 1: Summary of weighted sample of corn farms in Illinois Counties

County	N_i	n_i	$\bar{y}_{ij}$	min y_{ij}	max y_{ij}	$\bar{A}_{ij}$	min A_{ij}	max A_{ij}
De Witt	706	3	164.08	130.75	183.12	314.76	96.52	628.49
Logan	1023	5	164.88	125.66	201.21	558.90	424.90	860.45
Macon	917	1	167.10	167.10	167.10	1049.26	1049.26	1049.26
Marshall	708	4	200.52	189.60	208.48	487.82	281.33	682.75
Mason	871	4	178.41	146.35	203.95	509.69	245.52	1005.97
Mclean	1937	5	178.83	132.92	195.30	517.67	339.15	757.92
Menard	581	5	164.05	112.84	232.60	338.26	85.18	804.62
Peoria	1098	4	193.94	162.20	241.59	275.27	121.87	485.93
Stark	507	4	174.50	92.32	218.32	459.97	323.81	568.89
Tazewell	1245	5	175.43	147.75	193.62	518.62	197.49	889.58
Woodford	999	2	172.78	162.78	182.78	202.27	186.37	218.17
Champaign	1921	4	188.95	163.95	240.28	395.75	31.36	644.50
Ford	873	3	184.99	169.59	200.07	131.14	64.05	205.05
Iroquois	2041	2	164.95	136.55	193.36	540.53	493.27	587.79
Kankakee	1112	2	208.07	206.14	209.99	197.49	141.22	253.75
Livingstone	2022	1	142.08	142.08	142.08	163.68	163.68	163.68
Piatt	733	5	186.90	158.82	209.73	467.92	82.73	950.07
Vermillion	1608	4	171.93	128.81	206.32	382.91	235.74	702.99

range between 1 (as in Macon and Livingstone) and 5 (as in Logan, Mclean, Menard, Tazewell and Piatt). In each case the number of farms sampled is far less than the population in each county. Thus direct county averages are likely to be unreliable giving rise to a small area estimation problem. The sample corn yield ranges between 92 bu/acre and 242 bu/acre with county sample average ranging between 142 bu/acre and 209 bu/acre. The high sample mean yield in Kankakee and Marshall can be attributed to their relatively high rainfall during May to August (see table 8). The size of the farms sampled range from 31 acre to 1050 acre while the county average farm size ranges from 131 acre to 1050 acre [5].

4. Results and discussions

For illustrative and comparison purposes, table 6 reports expected residual losses (basis risk) on two representative farms in Champaign and Vermillion for all combinations of APH coverage (C_q), GRP coverage (C_z) and scale (S_p) based on the model with equal (B_{eq}) and unequal (B_{ueq}) weights.

The results show that for the farm in Champaign, almost no basis risk is triggered for APH coverages up to 80% for all GRP coverage-scale combinations. However, for an APH coverage of 85%, GRP coverage-scale combinations up to 85% trigger expected residual loss of 0.05 $bu/acre$. This loss is reduced to 0.02 $bu/acre$ by purchasing a GRP with 90% coverage and scale 0.9. Increasing the scale further reduces the loss proportionately. Notice that a GRP with 95% coverage completely tracks the loss at farm level. Thus a GRP with 95% coverage and scale 0.9 is the equivalent of an APH with 85% for an optimizing farmer. Similarly, an APH with 90% coverage triggers a higher expected residual loss compared to a 85% coverage with the same GRP coverage-scale combinations. These losses are gradually reduced with higher GRP coverage and scale but are not eliminated under the maximum GRP coverage and scale as is the case with an APH of 85%. A similar trend is observed with an APH coverage of 95% with different GRP policies. Basis risk increases at a decreasing rate for proportional increase in APH and GRP coverage for a given scale. For example, basis risk under an APH coverage of 90% and GRP coverage of 90% with a 1.5 scale is 380% higher than that under an APH and GRP with 85% coverage with the same scale. However, under an APH and GRP with 95% coverage and similar scale, basis risk is 120% higher than that observed under an 85% coverage for both plans with the same scale.

Unlike the farm in Champaign where there are some GRP coverage-scale combinations for which basis risk is zero, the farm in Vermillion appears to have no GRP policy which completely tracks the expected farm level loss the same as the APH policy. At the minimum coverage (70%) level, this farm has an expected residual loss of about 1 $bu/acre$ while at the maximum coverage (95%), the loss rises to 43 $bu/acre$. The expected residual losses tend to increase with an increase in APH coverage level for similar GRP coverage-scale policies. For a given APH and GRP coverage levels, the effect of scale on reducing basis risk is insignificant, ranging from about 0.2%-1% reduction in residual loss for 66.7% increase in scale from 0.9 to 1.5. For example, under a 75%-90%,80%-90%,85%-90% APH-GRP coverage combinations,

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

increasing the scale by 66.7% from 0.9 to 1.5 only leads to about 0.2% decrease in the expected residual loss. Whereas under a 75%-95%,80%-95%,85%-95% APH-GRP coverage combinations, increasing the scale by 66.7% from 0.9 to 1.5 leads to 0.8%, 0.9% and 1% decrease in the expected residual loss respectively. On the other hand, an increase in GRP coverage for the same APH coverage and scale (i.e., no scale effect) reduces the residual loss. However, this effect decreases with increase in APH coverage. For example, given a scale of 1.5, increasing the GRP coverage by 5% from 90% to 95% for APH policies each with 75%, 80%, 85%, 90% and 95% coverage reduces the expected residual loss by 3%, 2.6%, 2.2%, 1.8% and 1.3% respectively. The same effect under a scale of 0.9 results to slightly lower reductions (2.4%, 1.9%, 1.5%, 1% and 0.8%).

The results show that the efficiency of the expected residual loss estimates significantly increases with increase in APH coverage level. For example, at an APH coverage of 70% the t-statistics is about 0.76 while at an APH with with 95% coverage level the t-statistics is about 3.5 indicating that it becomes significant at a 5% level.

Except for De Witt county, there appears to be little or no differences in the estimates derived using data with equal and unequal weights even though estimates of expected county yield (as shown in table 2) from both models differ in some counties.

These results show potential wide variation in basis risk within farms in the same county as well as those across counties. Based on the results, a GRP policy with 95% coverage and scale 0.9 can replace an APH policy with 85% coverage for the farm in Champaign while the same GRP policy can only replace an APH policy with less than 70% coverage on the farm in Vermillion.

Table 7 presents results of expected residual loss by county for each APH coverage level averaged across all GRP coverage-scale combinations and farms under the two models. Similar to the results at farm level,we find a monotonic increase in expected residual loss with increase in APH coverage. Thus, the residual losses tend to be minimal at lower coverage levels. Note that the lowest coverage level is not necessarily the best in terms of minimizing residual losses since zero residual loss can be triggered at any coverage (although more likely at lower coverage) level which varies from one farm to another.

The results equally show wide variation in residual loss across counties for the same level of coverage. For example, Stark, Menard, De Witt, Logan, Woodford and Vermillion counties all exhibit high expected residual losses compared to Marshall, Mclean, Livingstone, Piatt, Macon, Champaign, Ford and Iroquois. This implies that expected farm level corn yield losses better correlate with expected county level loss in the later counties making area coverage crop insurance policies such as the GRP potentially more attractive in these counties. More specifically, in Marshall County a given GRP policy tracks all (0 expected residual loss) corn yield loss ($bu/acre$) the same as an APH with 80% coverage whereas, in Woodford County, a similar GRP policy tracks 12.23 $bu/acre$ less loss than an APH with 80% coverage. This means that GRP policies are likely to be more attractive to farmers in Marshall County than those in Woodford County. In addition and similar to results reported at farm level, the efficiency of the estimates increases with increase in the APH coverage.

Results from both models are similar at low coverage levels but slightly differ at high coverage levels. The magnitude of the estimated standard errors indicate that the estimates are more efficient with increase in the coverage level.

Table 2: Expected residual Corn Yield loss (basis risk) ($bu/acre$) on two farms

	Champaign		Vermillion			
C_q	C_z	S_p	$B_{eq}(s.e)$	$B_{ueq}(s.e)$	$B_{eq}(s.e)$	$B_{ueq}(s.e)$
0.70	0.70	0.9	0.00(0.00)	0.91(2.77)	0.00(0.00)	0.91(2.77)
”	”	”	”	”	”	”
0.70	0.85	1.5	0.00(0.00)	0.91(2.77)	0.00(0.00)	0.91(2.77)
0.70	0.90	0.9	0.00(0.00)	0.91(2.76)	0.00(0.00)	0.91(2.76)
0.70	0.90	1.5	0.00(0.00)	0.90(2.76)	0.00(0.00)	0.90(2.76)
0.70	0.95	0.9	0.00(0.00)	0.87(2.70)	0.00(0.00)	0.87(2.70)
0.70	0.95	1.5	0.00(0.00)	0.85(2.68)	0.00(0.00)	0.85(2.68)
0.75	0.70	0.9	0.00(0.00)	4.96(6.51)	0.00(0.00)	4.96(6.51)
”	”	”	”	”	”	”
0.75	0.80	1.5	0.00(0.00)	4.96(6.51)	0.00(0.00)	4.96(6.51)
0.75	0.85	0.9	0.00(0.00)	4.96(6.52)	0.00(0.00)	4.96(6.52)
0.75	0.85	1.5	0.00(0.00)	4.96(6.52)	0.00(0.00)	4.96(6.52)
0.75	0.90	0.9	0.00(0.00)	4.93(6.50)	0.00(0.00)	4.93(6.50)
0.75	0.90	1.5	0.00(0.00)	4.92(6.50)	0.00(0.00)	4.92(6.50)
0.75	0.95	0.9	0.00(0.00)	4.81(6.44)	0.00(0.00)	4.81(6.44)
0.75	0.95	1.5	0.00(0.00)	4.77(6.41)	0.00(0.00)	4.77(6.41)
0.80	0.70	0.9	0.00(0.06)	12.79(9.52)	0.00(0.06)	12.79(9.52)
”	”	”	”	”	”	”
0.80	0.85	1.5	0.00(0.06)	12.79(9.52)	0.00(0.06)	12.79(9.52)
0.80	0.90	0.9	0.00(0.00)	12.72(9.54)	0.00(0.00)	12.72(9.54)
0.80	0.90	1.5	0.00(0.00)	12.69(9.55)	0.00(0.00)	12.69(9.55)
0.80	0.95	0.9	0.00(0.00)	12.48(9.56)	0.00(0.00)	12.48(9.56)
0.80	0.95	1.5	0.00(0.00)	12.36(9.57)	0.00(0.00)	12.36(9.57)
0.85	0.70	0.9	0.05(0.72)	22.68(10.83)	0.05(0.72)	22.68(10.83)
”	”	”	”	”	”	”
0.85	0.85	1.5	0.05(0.71)	22.68(10.83)	0.05(0.71)	22.68(10.83)
0.85	0.90	0.9	0.02(0.31)	22.60(10.85)	0.02(0.31)	22.60(10.85)
0.85	0.90	1.5	0.00(0.06)	22.55(10.89)	0.00(0.06)	22.55(10.89)
0.85	0.95	0.9	0.00(0.00)	22.26(10.99)	0.00(0.00)	22.26(10.99)
0.85	0.95	1.5	0.00(0.00)	22.05(11.15)	0.00(0.00)	22.05(11.15)
0.90	0.70	0.9	0.36(2.01)	32.96(11.53)	0.36(2.01)	32.96(11.53)
”	”	”	”	”	”	”
0.90	0.85	1.5	0.32(1.89)	32.96(11.53)	0.32(1.89)	32.96(11.53)
0.90	0.90	0.9	0.24(1.43)	32.88(11.55)	0.24(1.43)	32.88(11.55)
0.90	0.90	1.5	0.21(1.23)	32.83(11.59)	0.21(1.23)	32.83(11.59)
0.90	0.95	0.9	0.09(0.76)	32.53(11.70)	0.09(0.76)	32.53(11.70)
0.90	0.95	1.5	0.06(0.64)	32.25(12.00)	0.06(0.64)	32.25(12.00)
0.95	0.70	0.9	1.51(4.42)	43.27(12.17)	1.51(4.42)	43.27(12.17)
”	”	”	”	”	”	”
0.95	0.85	1.5	1.47(4.28)	43.26(12.17)	1.47(4.28)	43.26(12.17)
0.95	0.90	0.9	1.27(3.77)	43.19(12.19)	1.27(3.77)	43.19(12.19)
0.95	0.90	1.5	1.16(3.55)	43.13(12.22)	1.16(3.55)	43.13(12.22)
0.95	0.95	0.9	0.72(2.69)	42.84(12.33)	0.72(2.69)	42.84(12.33)
0.95	0.95	1.5	0.53(2.26)	42.55(12.62)	0.53(2.26)	42.55(12.62)

5. Conclusion

This study attempted to quantify the magnitude of basis risk involved across corn farms in Illinois counties using pseudo-simulated yields with farm specific geospatial climate data.

Following Awondo et al. (2012) we used a two hierarchical Bayes small area estimation to address problems related to lack of representative sample, aggregation bias, properly accounting for spatial and temporal heterogeneity and uncertainty in parameter estimates.

We found wide variation in expected basis risk across farms within and between counties. Expected basis risk was found to sharply increase when comparing similar GRP policy with an APH policy with higher coverage level. However, changes in scale was found to have little effect on reducing basis risk between APH and GRP policy. The expected basis risk by county averaged over all scale for each coverage indicated relatively low basis risk in some counties, implying that farm level losses better correlate with losses at county level making these counties suitable for the introduction and expansion of area level policies such as GRP.

There is more room for research on developing small area estimates of expected yield and premiums such as dealing with outliers in the sample. Such is the case with Kankakee in the simulated sample (used in this study) which has very high observed yield and variance than all the other counties in the sample. Also, there is more room in prior development and sensitivity analysis. This include investigating the performance of this method with other crops and distributional forms. Extending the model to a time series framework will also be valuable.

Awondo, S. N., G. S. Datta, O. A. Ramirez, and E. G. Fonsah (2012, August 13-16.). Estimation of crop yield distribution and insurance premium using shrinkage estimator: A hierarchical bayes and small area estimation approach. Paper presented at Agricultural and Applied Economics Association Annual Meeting, Seattle, WA.

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*, 28–36.

Datta, G. S. and M. Ghosh (1991). Bayesian prediction in linear models: applications to small area estimation. *Annals of Statistics 19*, 1748–1770.

Datta, G. S., M. Ghosh, N. Nangia, and K. Natarajan (1996). *Estimation of median income of four-person families: A Bayesian approach, In Bayesian Analysis in Statistics and Econometrics*, pp. 129–140. Wiley. eds. D. A. Berry and K. M. Chaloner and J. K. Geweke.

Datta, G. S., M. Ghosh, and L. Waller (2000). *Hierarchical and empirical Bayes methods for environmental risk assessment, Handbook of Statistics, Bioenvironmental and Public Health Statistics* (18 ed.)., pp. 223–245. North-Holland. eds. P. K. Sen and C. R. Rao.

Datta, G. S., P. Lahiri, and T. Maiti (2002). Empirical bayes estimation of median income of four-person families by state using time series and cross-sectional data. *Journal of Statistical Planning and Inference 102*, 83–97.

Table 3: Expected residual corn yield loss(basis risk)($bu/acre$)by counties in Illinois based on data with unequal & equal weights

County	70%	75%	80%	85%	90%	95%
Unequal						
DeWitt	3.27(5.99)	6.44(10.23)	9.76(14.72)	13.14(19.32)	16.66(23.91)	21.04(28.07)
Logan	2.07(4.50)	5.42(8.44)	10.27(12.51)	16.10(16.60)	22.12(20.96)	28.19(25.51)
Macon	0.00(0.00)	0.01(0.26)	0.24(1.45)	1.15(3.71)	3.71(6.95)	8.07(10.61)
Marshall	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.04(0.50)	0.55(2.27)	2.64(5.50)
Mason	0.01(0.17)	0.20(1.19)	1.14(3.43)	3.05(6.57)	6.09(10.04)	11.15(13.29)
Mclean	0.09(0.95)	0.35(2.20)	0.87(3.93)	1.68(6.06)	2.81(8.49)	4.54(11.15)
Menard	4.16(8.42)	7.16(12.12)	11.03(15.95)	15.88(19.69)	21.76(23.17)	28.53(26.30)
Peoria	0.88(3.96)	2.16(7.00)	4.05(10.59)	6.63(14.54)	10.08(18.66)	14.44(22.87)
Stark	5.06(10.97)	7.05(14.27)	9.26(17.70)	11.73(21.21)	14.65(24.77)	18.09(28.37)
Tazewell	0.03(0.44)	0.34(1.78)	1.49(4.27)	3.77(7.59)	7.33(11.33)	11.87(15.23)
Woodford	3.08(7.06)	7.08(11.63)	12.23(16.57)	17.94(21.71)	23.90(26.99)	30.08(32.31)
Champaign	0.00(0.00)	0.00(0.05)	0.03(0.45)	0.31(1.69)	1.46(4.10)	4.09(7.42)
Ford	0.00(0.00)	0.02(0.37)	0.23(1.54)	1.07(3.83)	3.05(7.20)	6.16(11.17)
Iroquois	0.07(1.00)	0.24(1.99)	0.68(3.49)	1.39(5.48)	2.47(7.81)	4.14(10.45)
Kankakee	0.01(0.25)	0.14(1.26)	0.74(3.48)	2.35(6.82)	5.71(11.08)	11.05(15.69)
Livingstone	0.03(0.44)	0.22(1.54)	0.76(3.29)	1.82(5.66)	3.47(8.50)	5.56(11.64)
Piatt	0.00(0.14)	0.03(0.60)	0.14(1.39)	0.54(2.87)	1.38(5.04)	2.90(7.79)
Vermillion	2.31(6.59)	4.98(9.99)	8.95(13.77)	13.56(17.86)	18.35(22.24)	23.24(26.78)
Equal						
DeWitt	3.63(6.23)	6.97(10.60)	10.34(15.20)	13.72(19.88)	17.25(24.51)	21.72(28.64)
Logan	2.07(4.50)	5.42(8.44)	10.27(12.51)	16.10(16.60)	22.12(20.96)	28.19(25.51)
Macon	0.00(0.00)	0.01(0.26)	0.24(1.45)	1.15(3.71)	3.71(6.95)	8.07(10.61)
Marshall	0.00(0.00)	0.00(0.00)	0.00(0.00)	0.04(0.50)	0.55(2.27)	2.64(5.50)
Mason	0.01(0.17)	0.20(1.19)	1.14(3.43)	3.05(6.57)	6.09(10.04)	11.15(13.29)
Mclean	0.09(0.95)	0.35(2.19)	0.86(3.92)	1.68(6.05)	2.80(8.47)	4.52(11.13)
Menard	4.13(8.36)	7.14(12.06)	11.01(15.90)	15.85(19.64)	21.72(23.13)	28.48(26.27)
Peoria	0.89(3.96)	2.16(7.00)	4.05(10.60)	6.62(14.54)	10.07(18.66)	14.43(22.87)
Stark	5.16(11.09)	7.17(14.41)	9.38(17.85)	11.86(21.37)	14.79(24.93)	18.24(28.52)
Tazewell	0.03(0.44)	0.35(1.78)	1.50(4.28)	3.79(7.61)	7.35(11.35)	11.90(15.25)
Woodford	3.07(7.04)	7.07(11.62)	12.22(16.55)	17.93(21.69)	23.89(26.98)	30.07(32.30)
Champaign	0.00(0.00)	0.00(0.05)	0.03(0.45)	0.31(1.70)	1.49(4.14)	4.21(7.48)
Ford	0.00(0.00)	0.02(0.37)	0.23(1.54)	1.08(3.83)	3.06(7.21)	6.20(11.18)
Iroquois	0.07(1.00)	0.24(1.99)	0.68(3.49)	1.40(5.48)	2.48(7.81)	4.15(10.45)
Kankakee	0.00(0.19)	0.12(1.13)	0.67(3.28)	2.20(6.54)	5.43(10.75)	10.61(15.36)
Livingstone	0.03(0.44)	0.22(1.54)	0.76(3.29)	1.82(5.66)	3.47(8.50)	5.56(11.64)
Piatt	0.00(0.14)	0.03(0.60)	0.15(1.40)	0.56(2.92)	1.46(5.15)	3.07(7.95)
Vermillion	2.31(6.59)	4.98(9.99)	8.94(13.77)	13.55(17.86)	18.34(22.24)	23.22(26.78)

Datta, G. S., P. Lahiri, 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*, 1074–1082.

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*, 269–277.

Gelfand, A. E. and A. F. M. Smith (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 398–409.

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*, 1423–1431.

Ghosh, M. and J. N. K. Rao (1994). Small area estimation: an appraisal. *Statistical Science 9*, 55–93.

Prasad, N. G. N. and J. N. K. Rao (1990). The estimation of mean squared error of small area estimators. *J. Amer. Statist. Assoc. 85*, 163–171.

Prasad, N. G. N. and J. N. K. Rao (1999). On robust small area estimation using a simple random effects model. *Survey Methodology 25*, 67–72.

Ramirez, O. A., T. U. McDonald, and C. A. Carpio (2010). A flexible parametric family for the modeling and simulation of yield distributions. *Journal of Agricultural and Applied Economics 42*(2), 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.

Schlenker, W. and M. Roberts (2006). The impact of global warming on u.s. agriculture: An econometric analysis of optimal growing conditions. *Review of Economics and Statistics 88*(1), 113–125.

Tannura, M. A., S. H. Irwin, and D. L. Good (2008, February). Weather, technology, and corn and soybean yields in the u.s. corn belt. marketing and outlook research report.

Thompson, L. M. (1988). Effects of changes in climate and weather variability on the yields of corn and soybeans. *Journal of Production Agriculture 1*, 20–27.

Torabi, M. (2012). Hierarchical bayes estimation of spatial statistics. *Journal of Statistical Planning and Inference 142*, 358–365.

USDA-NASS (2009). Area frame design for agricultural surveys. Research and Development Division, Washington DC.

Wang, Y., B. J. Barnett, K. H. Coble, and A. Harri (2012, February 4-7). An examination of how the level of yield aggregation impacts the effectiveness of area-based insurance designs. Paper presented at the annual meeting of the Southern Agricultural Economics Association, Birmingham, Alabama.

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*, 197–208.

Table 4: Summary of monthly cumulative precipitation (inches) and average temperature (F),May-August for Illinois counties

County	P_5	P_6	P_7	P_8	T_5	T_6	T_7	T_8
De Witt	4.64	5.13	1.79	1.23	61.98	72.50	79.66	74.06
Logan	4.56	5.46	1.83	1.09	62.11	72.57	79.81	74.32
Macon	4.51	6.12	1.96	0.82	62.55	73.07	80.25	74.98
Marshall	5.39	4.85	2.55	2.60	61.37	71.64	79.09	73.17
Mason	4.57	5.45	1.87	1.12	61.97	72.38	79.72	74.18
Mclean	4.89	4.87	2.09	1.64	61.62	72.06	79.45	73.74
Menard	4.49	5.80	1.84	0.94	62.33	72.79	80.02	74.70
Peoria	5.28	4.91	2.31	2.03	61.43	71.72	79.08	73.32
Stark	5.56	4.99	2.70	2.68	61.35	71.44	79.12	73.28
Tazewell	4.85	4.91	2.04	1.57	61.64	72.07	79.49	73.78
Woodford	5.44	5.11	2.47	2.03	61.37	71.58	79.16	73.37
Champaign	4.62	5.41	1.85	1.16	62.04	72.47	79.74	74.24
Ford	5.09	4.86	2.28	1.89	61.51	71.84	79.29	73.53
Iroquois	5.37	4.84	2.46	2.21	61.41	71.65	79.11	73.30
Kankakee	5.63	5.15	2.51	3.26	60.84	71.09	78.66	72.84
Livingstone	5.50	4.94	2.51	2.39	61.27	71.49	79.03	73.15
Piatt	4.56	5.77	1.85	0.99	62.26	72.72	79.97	74.59
Vermillion	4.64	5.37	1.94	1.22	61.92	72.31	79.69	74.15

Table 5: Summary posterior distribution

Parameter	mean	sd	MC error	2.5%	median	97.5%
b0	-36.39	322.5	10.02	-654.5	-39.01	587.3
P_5	71.44	132.7	3.89	-190.4	72.02	333.8
P_6	176.0	69.94	2.216	41.13	175.3	317.8
P_7	67.17	169.5	5.177	-255.6	67.07	391.3
P_8	-9.287	156.8	4.774	-320.2	-6.706	289.5
P_5^2	-5.457	11.14	0.3265	-27.44	-5.176	16.13
P_6^2	-14.05	7.255	0.238	-28.18	-13.81	-0.2592
P_7^2	-33.51	41.37	1.305	-115.2	-32.73	45.86
P_8^2	9.827	37.14	1.142	-59.71	8.814	83.33
T_5	27.08	29.47	0.8278	-30.82	28.23	85.57
T_6	-10.19	39.71	1.176	-89.13	-9.627	69.68
T_7	-2.153	34.09	1.148	-69.08	-1.801	65.92
T_8	-17.59	43.56	1.411	-103.2	-16.99	66.3
σ_v^2	15.09	3.939	0.1199	8.685	14.76	22.97
σ_e^2	0.002161	3.031E-4	7.372E-6	0.001587	0.002162	0.002797
deviance	616.5	7.287	0.2125	604.8	615.8	632.7