Can Crop Insurance Premiums be Reliably Estimated?

by

Octavio A. Ramirez
Professor and Head, Department of Agricultural and Applied Economics,
University of Georgia, Athens, GA 30602, oramirez@uga.edu

Carlos E. Carpio
Assistant Professor, Department of Applied Economics and Statistics,
Clemson University, Clemson, SC 29634, ccarpio@clemson.edu

and

Rod Rejesus
Assistant Professor, Department of Agricultural and Resource Economics,
North Carolina State University, Raleigh NC 27695, rod_rejesus@ncsu.edu

Paper Accepted for Presentation at the Annual Meeting of the *American Agricultural Economics Association* Milwaukee, WI, July 26-28, 2009

Can Crop Insurance Premiums be Reliably Estimated?

By far the most popular risk management tool used by U.S. crop producers is crop insurance. In 2008, the U.S. crop insurance program insured more than 270 million acres, representing more than $89 billion in value, with total premium of over $9 billion. This level of coverage has largely been obtained due to greatly increased premium subsidies that encourage producers to purchase higher coverage levels (Glauber, 2004). Not unexpectedly, there has been a corresponding increase in expected annual costs of delivering the program to over $3 billion, compared to just one-third of that cost in 1990 (Glauber, 2004; Worth, 2008).

In light of its role as a major safety-net for crop producers and the cost of delivering this program, significant research efforts have been conducted to improve premium rating procedures and the actuarial performance of the program (Glauber, 2004). Premium rating procedures that more accurately reflect producer risks mitigate adverse selection and moral hazard problems. More accurate rating will also assure that the program meets the risk management needs of producers at a reasonable cost to the nation's taxpayers.

Despite the improvement in the program's aggregate actuarial performance (i.e. relative to the performance in the 1980s and 1990s), there is still a common perception that the program may not be properly rated because there are large regional disparities in loss experience across crops and regions (Glauber, 2004, Babcock, 2008). For the 1981-2003 period, for example, the Midwest in general have received less indemnities than premiums paid (i.e. loss ratio = indemnity/premium < 1), relative to the Great Plains and the Southeast (loss ratio > 1.5). The regional disparities in loss performance suggest program benefits are not being equitably distributed (i.e. Midwest farmers are paying more into the program than they have received benefits in return). The most common reasons put forward for the regional inequities in loss performance are political influences and systematic misratings (Glauber, 2004; Goodwin and Vado, 2007; Babcock, 2008).

Although an extensive and highly relevant body work on crop insurance programs and rating has been published in the agricultural economics literature, to date, the question of how accurately can actual crop insurance premiums be estimated via the RMA procedures and other proposed statistical methods remains unanswered. The reason for this gap in the literature might be that, in principle, the "true" (i.e. 100% actuarially correct) premiums that would need to be charged in a particular case are unknown, making any comparison between estimated and true rates unfeasible.

This research makes such a comparison possible through simulation methods. Specifically, yield distributions are estimated on the basis of a comprehensive farm-level dataset in the U.S., using recently developed parametric modeling procedures that are sufficiently flexible to accommodate the variety of distributional shapes that might be encountered in practice (Ramirez and McDonald 2007). Although these estimates will not be totally accurate, they should sufficiently resemble the true underlying distributional shapes to make the analyses realistic.

The objective of this paper, therefore, is to determine the accuracy of various rating methods (at the county and individual level) by comparing the estimated rates to "true" rates using empirically-grounded simulation procedures that takes into account common data availability constraints. Presumably, if one has a rich data set with a large number of farmers each having a very long time-series of yield and insurance information (i.e. indemnities and liabilities), one can come close to accurately estimating the "true" premium rate. However, this type of data is rarely available and, therefore, it is important to assess different premium rate-setting methods in the context of limited data availability (i.e. limited number of farms in the sample and limited number of years for each farm).

Undertaking this analysis provides important implications as to what type of rating procedures perform well under different data environments and assumptions. The current Risk Management Agency (RMA) rating approach is primarily a "loss-cost" approach that relies primarily on historical liability and indemnity data in setting the rates. A "yield distribution" simulation approach to setting premiums, where assumed yield distributions are used to simulate losses, is another method typically used to set rates. The empirical evidence on these two approaches seem to suggest that rates from "yield distribution" based approaches tend to be lower than the rates from a "loss-cost" based approach (see Deng, Barnett, and Vedenov, 2007). The present study adds more to this literature by providing insights on how the number of farms and the number of years (for each farm) in the data set affects the rating performance of the two common approaches to premium rate-setting. Would having more number of farms in the data set improve accuracy more than adding more years for each farm? Furthermore, the effects of different factors/assumptions (i.e. correlation of yields across farms, non-normality assumptions, county versus individual rates) on rate accuracy are investigated as well and results from this analysis can also help guide researchers in improving rate-making procedures.

Methods and Procedures

The data for this study was obtained from the University of Illinois Endowment Farms database, which includes 26 corn farms located in twelve counties across that State. Data were available from 1959 to 2003, with sample sizes ranging from 20 to 45 observations. Ramirez and McDonald (2006) used this data to estimate models for those 26 yield distributions that are as accurate (i.e. realistic) as possible. To this effect, they used a system of probability distributions that has sufficient flexibility to parametrically model any empirically possible distributional shape with a high level of accuracy. This system, which is composed of the S_U and the S_B families (Johnson, 1949), can accommodate any mean-variance-skewness-kurtosis (MVSK) combination that might be encountered in practice, and allows for the testing of the null hypothesis of yield normality.

An advantage of using Ramirez and McDonald (2006) results for this research is that they identify a variety of distributional shapes that span over a substantial area of the theoretically feasible skewness-kurtosis (SK) space. A thoughtfully selected subset of these 26 models should, therefore, be representative of the breadth of distributional shapes that could possibly be encountered in practice.

In their analyses, Ramirez and McDonald (2006) estimate normal, S_U and S_B models for each of the 26 yield series. They then conduct likelihood ratio tests which reject the normality hypothesis in 20 of the 26 cases (α=0.10). All estimated non-normal distributions are found to be left-skewed, which is consistent with previous literature (Nelson and Preckel, 1989; Taylor, 1990; Ramirez, 1997; Ker and Coble, 2003; Harri, Coble, Erdem and Knight, 2005). Out of the 20 cases that are classified as non-normal, the S_B models exhibit the highest maximum log likelihood function values in 14 cases and the S_U models in six cases.

Five of the 26 distributions estimated by Ramirez and McDonald (2006) are chosen for the purposes of this research. These include a normal, two S_U's and two S_B's. The S_U's and S_B's are selected to have: a) Low skewness and relatively high positive kurtosis (S_{UA}), b) Low skewness and relatively high negative kurtosis (S_{BA}), c) Moderately negative skewness and positive kurtosis (S_{UB}), and d) Relatively high skewness and kurtosis (S_{BB}). That is, they are representative of the SK spectrum of the 26 distributional shapes identified by Ramirez and McDonald (2006) associated with empirical farm-level yield data. For the purposes of this research, these are assumed to be the distributional shapes underlying five typical yield data generating processes.

The research also requires simulating draws from the S_U and S_B distributions. The simulation formulas (Ramirez, McDonald 2006) are:

(1) $SS_U = M + \{\sigma[\sinh(\theta\{Z + \mu\}) - F_{SU}] \div \theta\sqrt{G_{SU}}\}$ for the S_U,

(2) $SS_{SB} = M + \{\sigma \exp(\theta[Z - \mu]) \div \sqrt{G_{SB}}[1 + \exp(\theta[Z - \mu]) - F_{SB}]\}$ for the S_B, and

(3) $SS_N = \sigma Z + M$ for the normal distribution,

where M and σ are the mean and variance, μ and θ are the shape parameters, Z is a draw from a standard normal, and F_{SU}, G_{SU}, F_{SB}, and G_{SB} are lengthy exponential and trigonometric functions of μ and θ (equations 7 and 9 in Ramirez and McDonald 2006).

The next step is to use the previous formulas to simulate data from the five selected distributions (Normal, S_{UA} S_{UB} S_{BA} S_{BB}). The skewness and kurtosis parameters estimated by Ramirez and McDonald (2006) are used in the four non-normal cases. The means and variances of the simulated distributions, however, are adjusted to meet a key objective of the research. Specifically, NF sets of mean and standard deviation parameters are assumed to be drawn from uniform distributions with ranges of 140 to 180 bushes/acre and 25 to 35 bushels per acre, respectively. These are consistent with the range of mean and variances reported by Ramirez and McDonald (2006) for their 26 estimated corn yield distributions.

The reason for this framework is to explore a hypothetical situation where one observes yields from a number of farms (NF) within the same region or county, which have different mean and variances but the same distributional shape (i.e. SK) characteristics. The fact that the distributional shapes used in this evaluation are empirically motivated, i.e., derived from parametric models that have been estimated on the basis of actual yield data, enhances the credibility of the analysis.

The exact actuarially fair crop insurance premiums corresponding to each of the “assumed” yield distributions for the typical 65% coverage level are then computed on the basis of simulated yield samples of size 100,000, using standard procedures (i.e. this is like having 100,000 years of data for one of the NF farms). Specifically, each of the 100,000 simulated yield values (Y_i) is compared with

0.65 times the mean of the entire sample ($\bar{Y}$). If the actual yield value is lower than $0.65x\bar{Y}$, the difference ($0.65x\bar{Y}$ -Y_i) is multiplied by the assumed price guarantee (Pg=$2.2/bushel in this case), otherwise the observation is discarded. Then sum of all the non-discarded values divided by 100,000 is the expected indemnity associated with that yield distribution and, therefore, the actuarially fair premium to be charged to that farm.

For the purposes of this research, these are considered to be the true (i.e. correct) premiums corresponding to the "assumed" distributions. Since there are NF "assumed" mean and variance sets, this process is repeated NF times for each of the five selected distributions, resulting in the NF true premiums corresponding to each of the "farms" in the "county." In addition, runs for three different NF values (100, 50 and 25) are conducted. Thus, the final output is 100, 50 and 25 sets of true premiums for each of the distributions in the analysis.

The next step is to attempt to estimate premium rates under realistic field conditions. To this effect, random samples of size of SS = 10, 25 and 50 are simulated using the same NF sets of mean and variance parameters assumed in the computation of the true premiums as well as the estimated shape parameters corresponding to each particular distribution. Such samples are generated for NF = 100, 50 and 25 and correlation coefficients of CC = 0 and 0.5.

The unit of analysis is a particular NF-SS-CC combination. Therefore, for each distribution, there are 3x3x2=18 units of analysis. Given the data availability constraints we impose above, the best situation for estimating county and farm-level rates would be to have data on NF=100 farms with SS=50 observations for each and no correlation across farm yields (unit of analysis 100-50-0). The worst case scenario would then be having data on only NF=25 farms with SS=10 observations for each and a 0.5 correlation across them (unit of analysis 25-10-0.5). The remaining combinations span the spectrum between these two scenarios.

The next step is to use the same distribution utilized to simulate the previously discussed datasets to then jointly estimate the yield distributions of all NF "farms" within each unit of analysis (i.e. data availabilty constraints imposed). Three alternative joint probability distribution models are specified and estimated: one with separate means and variances for each farm (as in the data-generating process) (**M1**), one with a constant mean and variance (**M2**), one with different means but a constant variance for all farms (**M3**). The estimated models are then used to simulate yield draws (n=100,000 each) and compute actuarially fair premiums following the previously described procedure. In addition, two RMA-like rate computation protocols and a method assuming a normal distribution are also applied to estimate the premiums and they are explained in the next section.

RMA-like procedures

The RMA-like procedures used in this study are based on the individual empirical premiums implied by the Actual Production History (APH) yield data. The empirical premium for coverage at the 0.65 x 100% of the APH level of coverage is given by:

(4) $$ER_i = [\sum_{t=1}^{n} P_g Y_i^* / SS]$$

where $Y_i^* = \begin{cases} 0.65 - Y_{it} & if \quad Y_{it} < 0.65 APH_{it} \\ 0 & otherwise \end{cases}$,

where ER_i is ith's farmer empirical APH based premium, 0.65 is the level of coverage as explained previously, P_g is the guaranteed price which is assumed constant, SS is the number of years or sample size, t is the subscript to denote the year, APH_{it} and Y_{it} are the RMA APH approved and observed yield data for the *i* farmer in year *t*, respectively. Equation (4) is similar to empirical rate presented in Skees and Reed (1986) and Goodwin (1994); however, the mean yield (μ) in their equations is replaced by APH_{it}. In practice, the APH_{it} is the average of at least 4 years of individual yield data. At the beginning of the historical period, when a farmer just enters into the insurance program, the RMA assigns a transitional yield (t-yields) based on the county average and therefore the RMA APH approved yields are not equal to the farm observed average yields during the first four years of "history." Hence, we simulated the APH_{it} yields as follows: APH_{i1} for all i's was the average yield of a different batch of yield simulations for each unit of analysis. This is meant to simulate the average group yield from previous years (t-yield). APH_{i2} will be the actual t=1 yield realization (drawn from the simulation) + (t-yield x 3 years) and this value divided by four (since 4 years of data). And so on until t=4, when 4 years of actual yield data are available to calculate the individual APH_{i4}. Subsequently, APH_{it} are updated as more yield observations become available.

The empirical rate calculation based on the APH yields (**M4**) is actually not used by the RMA but it is included as one of the RMA-like procedures for two reasons. First, its calculation resembles the procedures used to calculate the group (county) level insurance rates and second it uses RMA APH approved yields. Moreover, these premium rates are used to simulate historical group indemnities (losses) and liabilities which constitute the main ingredients of the current RMA procedure. From a statistical perspective, M4 is a non-parametric procedure for the calculation of the premium rates that uses APH yields instead of the observed yields.

The next method used in the simulation is the "exponential" based RMA type of procedure (**M5**). This procedure is based on the main equation underlying the RMA's current ratemaking procedure (Milliman and Robertson, 2000):

$$RMA\_PR_i = P_g \text{x} 0.65 \text{x} APH_{iSS} \text{x} CPR \text{x} \left(\frac{APH_{iSS}}{Yavc}\right)^{Exponential} , \quad (5)$$

where RMA_PR_i is the RMA exponential based premium rate, P_g is the guaranteed price, *CPR* is the county rate, *Exponential* is just an exponent that is usually less than -1, and APH_{iSS} and are $Yavc$ the farm *i* APH yield and county average, respectively (Milliman and Robertson, 2000). Both APH_{iSS} and

$Yavc$ are calculated using the entire sample of simulated observations (SS).[1] This is a simplified version of the equation used by the RMA, but includes all the elements that are central for our analysis.[2] The logic underlying the equation is that individual farmers' premiums rates can be found using as the baseline the county rate. The *Exponential* is used so that farmers with above area average yields pay lower premiums and farmers with below area average yields pay higher insurance premiums (Knight, 2000).

The calculation of *CR* was based on the simulated farm level (*i* subscript) indemnities and liabilities for each sample (*t* subscript). The simulated indemnity, liability and CR in year t for the NF group of farms are:

(6) $$Indemnity_t = \sum_{t=1}^{NF} P_g Y_i^*$$

(7) $$Liability_t = \sum_{t=1}^{NF} P_g 0.65APH_i = P_g \text{x} 0.65 \text{x} NF \text{x} Yavc$$

(8) $$CR_t = \frac{Indemnity_{\ t}}{Liability_{\ t}}$$

Hence, the simulated CR using the SS observations in the sample is:

(9) $$CR_t = \frac{1}{SS}\sum_{t=1}^{SS} \frac{\sum_{i=1}^{NF} Y_i^*}{0.65\text{x}NF\text{x}Yavc} = \frac{\overline{ER}}{0.65Yavc}$$

Where Y_i^* was defined in (4) and $\overline{ER}$ is the average empirical premium rate across all farms.

The *Exponential* is estimated then using non-linear least squares (NLLS) with the following regression model:[3]

(10) $$ER_i = P_g \text{x} 0.65 \text{x} APH_{iSS} \text{x} CPR \text{x} (\frac{APH_{iSS}}{Yavc})^{Exponential} + \varepsilon_i,$$

---

[1] In the RMA literature the ratio $\frac{APH_{iSS}}{Yavc}$ is usually called "*yield span*".

[2] The RMA procedure includes a myriad of other adjustments including caps for annual change of premiums levels, adjusting losses and exposures to a common coverage level, etc. (Milliman and Robertson, 2000).

[3] The actual method used by the RMA to calculate the exponential is not publically available. The only RMA document where exponentials are calculated is Knights' (2000) examination of yield span adjustments. This author uses an equation similar to (4) but the Heckman two-step procedure is used instead of NLLS.

where ε_i is the error term. NLLS was chosen given the presence of several ER_i estimates were equal to zero that preclude us from linearizing the expression using logs. The nonlinear censored regression approach proposed by (Stute, 1999) was also used in the preliminary stages of the study but the results obtained were not very different than those obtained with the NLLS estimation procedure, hence because of space limitations we excluded it from the simulations.

Premiums Rates Based on the Normal Distribution

The final method used in the simulations calculates individual premiums rates assuming a normal distribution (**M6**) (Skees and Reed, 1986; Goodwin, 1994):

(11) $$NR_i = P_g\left[\Phi\left(\frac{0.65\bar{Y}_i - \bar{Y}_i}{\hat{\sigma}_i}\right)(0.65\bar{Y}_i - \bar{Y}_i) + \varphi\left(\frac{0.65\bar{Y}_i - \bar{Y}_i}{\hat{\sigma}_i}\right)\hat{\sigma}_i\right],$$

where NR_i is the premium rate calculated assuming a normal distribution, Φ is the standard normal cdf, φ is the standard normal pdf, and $\bar{Y}_i$ and $\hat{\sigma}_i$ are the estimated mean and standard error of farmer' *i* yields calculated using the SS observations. NR_i was included in the calculations to assess the impact of incorrect distributional assumptions on the premium rate calculations.

In summary, 6 premium estimation procedures are considered. These methods can be grouped into three categories:

1) Premium procedures based on statistical yield models that attempt to estimate the true yield distribution or simplified versions of it: M1 (estimates means and variances for all farms in the group), M2 (estimates only one mean and one variance), and M3 (estimates different means and one variance for the group of farms).
2) Methods based on historical losses: M4 (uses observed individual indemnities calculated using RMA APH approved yields) and M5 (RMA-like procedure based on observed group indemnities and liabilities).
3) Premium rates calculated assuming a normal distribution.

Comparison Criteria

The premiums estimated through the six procedures (M1, M2,…,M6) are then compared with the true (i.e. correct) rates in order to ascertain how accurately crop insurance premiums can be estimated under a variety of empirically plausible data availability conditions, model specifications strategies, and rate computation procedures. The statistics used to compare the estimated premiums with the true premiums are:

1) Mean absolute error of the estimated premiums at the farm level (Farm-MAD):

(12) $$\left(\frac{1}{50xNF}\right)\sum_{j=1}^{NF}\sum_{r=1}^{50}\left|\hat{P}_{rj} - Ptrue_{rj}\right| ,$$

where $\hat{P}_{rj}$ is the estimated premium for farm *j* in run *r*, and $Ptrue_{rj}$ is the corresponding true premium value. The Farm-MAD statistic measures the accuracy of the estimated premiums at the farm level.

2) Mean absolute error of the average premiums at the group level (Group-MAD):

(13) $$\left(\frac{1}{50}\right)\sum_{j=1}^{50}\left|\left(\frac{1}{NF}\right)\sum_{j=1}^{NF}\hat{P}_{rj} - \left(\frac{1}{NF}\right)\sum_{j-1}^{NF}Ptrue_{rj}\right|.$$

3) The difference between the average estimated premium and the average true premium at the group level (Group-Bias):

(14) $$\left(\frac{1}{50xNF}\right)\sum_{j=1}^{50}\sum_{j=1}^{NF}\hat{P}_{rj} - \left(\frac{1}{50xNF}\right)\sum_{j=1}^{50}\sum_{j=1}^{NF}Ptrue_{rj} .$$

Both, the Group-MAD and the Group-Bias statistics measure the accuracy of the group average estimated premiums. As explained in a previous section, the RMA uses the county (i.e., a group of farmers) as the basic ratemaking unit; hence it is important to analyze the precision of the estimated average group premiums relative to their corresponding true values.

Effect of SS, NF, C and premium computation procedures on accuracy of premiums

The relationships between the measures of error/bias (i.e. Group-Bias, Farm-MAD and Group-MAD) and the rate computation procedures, SS, NF and CC were examined using regression models of the following form:

(15)
$$y = \beta_0 + \beta_{SS} \ln(SS) + \beta_{NF} \ln(NF) + \beta_{CC} CC + \sum_{j=2}^{6} \beta_{M_j} D_{M_j} + \sum_{j=2}^{6} \beta_{M_j,SS} D_{M_j} \ln(SS) + \sum_{j=2}^{6} \beta_{M_j,NF} D_{M_j} \ln(NF) + \sum_{j=2}^{6} \beta_{M_j,CC} D_{M_j} CC + \varepsilon.$$

The regression models used as dependent variables (y): the Group-Bias, Farm-MAD and Group-MAD. As explanatory variables the model included the natural log (ln) of SS, the ln of NF, CC, dummy variables for each estimation procedure ($D_{M_j}, j = 2,3,4,5$) and simple interactions between ln(SS), ln(NF), CC and the method. The β's are parameters corresponding to the explanatory variables.

To avoid perfect multicollinearity the dummy variable corresponding to estimation procedure 1 (D_{M_1}) was dropped from the models. Separate models using OLS were estimated for each of the five distributions, resulting in a total of 15 regression models (three measures of accuracy times 5 distributions). Standard errors were estimated using White heteroskedastic consistent covariance matrix.

The use of the dummy variables in model (7) allows for the estimation and testing of premium-estimation-procedure specific intercept and slope coefficients. Hence model (7) was used as the baseline for the estimation of models restricting some of the intercepts and slope parameters to be equal across models and/or equal to zero. F tests were conducted to verify that the set of parameter restrictions imposed in each of the final models were statistically valid. The parameter estimates of the restricted and their corresponding covariance matrix were then used to estimate the intercepts ($\beta_0 + \beta_{M_j}$) and slope coefficients ($(\beta_{SS}+\beta_{M_j,SS}), (\beta_{NF} + \beta_{M_j,NF}), and\ (\beta_{CC} + \beta_{M_j,CC})$) for each premium estimation method M_j.

Results

Simulation Results

Table 1 shows an example of the results obtained from the simulation experiments. The figures displayed correspond to the Farm-MAD simulation results assuming a normal distribution. A total of 108 Farm-MAD values were calculated corresponding to each particular combination of SS-NF-CC and premium estimation procedures. To facilitate comparison of the results across different distributions, all the Farm-MAD values (as well the other accuracy measures) are calculated relative to normalized average true premiums equal to 10. For example, the 6.06 figure displayed in the first row and first column of the table (gray area) corresponds to the Farm-MAD of the premiums estimated using M1, with 25 farm units (NF=25) in the group, a sample size of 10 observations per farm (SS=10), 0 correlation between farm yields in the group (CC=0), when the underlying true distribution is normal. Since the normalized true premium equals 10, the 6.06 value indicates that the farm level premiums estimated using M1 are, on average, 61% above or below their true values. Tables similar to Table 1 were constructed for all the five assumed distributions and the three measures of accuracy (available from the authors upon request) and used to estimate the regression models of the form shown equation (15).

Regression Analysis Results

Table 2 summarizes the regression analyses results for the Farm-MAD variable under the five assumed yield distributions. Except for the Farm-MAD regression model for the SBN distribution, all the other regression models had R^2 values higher than 0.93, suggesting that most of the variability in the Farm-MAD variable is explained by SS, NF and CC. However, the sign and magnitude of the parameter estimates suggest that the effect of these variables differs across distributions and premium estimation procedures. The only result that was highly consistent across all methods and distributions is the negative effect of the sample size which indicates that sample size, as expected, improves the accuracy of the farm level premium estimates. The number of farms (NF) did not have a statistically significant effect in about 60% of the cases and in the remaining 40% of cases where it was significant its effect was either positive or negative. Regarding the effect of CC on the Farm-MAD, in most cases a higher CC increases Farm-MAD.

Group-MAD regression analyses are shown next in table 3. For M1, M2 and M3, the overall effect of SS over Group-MAD was similar to the effect of these variables over Farm-MAD (i.e., SS reduces Group-MADs). For these three methods, NF is also found to improve (i.e. reduce) the Group-MADs in most of the cases. On the other hand, the effect of SS and NF in the performance of M4, M5 and M6 is less conclusive. With regard to the effect of CC, across most methods and distributions CC has a significant and positive effect, except for M1, M2 and M3 in distribution SBA for which the effect is negative (somehow consistent with the result observed in the corresponding Farm-MAD regression model).

Table 4 displays regression analysis results for Group-Bias. Given the fact that the bias can be positive or negative, the interpretation of the parameter estimate is done in relation to the value of the intercept which in theory reflects the bias of the procedure when SS, NF and CC equal zero. Similarly to the previous two regression models, an increase in SS, in the majority of cases, improves the accuracy (i.e., reduces the bias) of the estimated premiums. The effects of NF and CC are less conclusive. The only method where the statistically significant parameters had the expected effect across all distributions was M1.

Taken as a whole, the analysis of the regression results across all distributions suggests the following:

a) The sample size (SS) is the only variable that consistently improves the accuracy of the estimated premiums across most of the methods.

b) The effect of SS and NF on Farm-MAD and Group-MAD for M1, M2 and M3 are generally consistent with previous expectations and tended to improve the accuracy of the estimation as compared to the effect of these variables for M4, M5 and M6. This result might have to do with the fact that M1, M2 and M3 use observed yield data. On the other hand, M4, M5 and M6 use APH yields which as explained in the methods section are not necessarily equal to the observed yields.

c) The effect of SS on Farm-MAD and Group-MAD for M1, M2 and M3 is always higher than the effect of NF. In other words, each additional observation improves more the MAD of the estimated premiums than each additional farm added to the group.

d) The effect of SS and NF on Farm-MAD and Group-MAD for M1 and M3, as measured by the estimated parameter values, is always equal or higher than the effect of these variables for M2. This result might have to do with the fact that M1 and M3 estimate models that are closer to the true model than the model estimated using M2.

e) Regarding the Group-Bias, the only two methods where SS and NF had the expected bias-reducing effect were M1 and M3, which as explained previously are the methods that most closely resemble the "true underlying model."

f) The CC in most cases has a negative effect on the accuracy of the estimated premiums. This effect was more clearly observed in the Group-MAD regression models where 27 out of the 30 parameters corresponding to this variable were positive and statistically significant. In the other two sets of regression models (Farm-MAD and Group-Bias), the CC variable reduces accuracy of the estimation in about 50% of the cases with the remaining 50% of cases almost equally split between being no-significant or improving the accuracy of the estimated premiums.

Comparison using Predicted Accuracy Measures

Whereas the regression analysis allowed us to analyze the general direction of the effects of NF, SS and CC on the accuracy of the estimated premiums, a better picture of the magnitude and economic importance of the effects of these variables can be obtained by looking at the predicted values of the accuracy measures, especially since both NF and SS are nonlinearly related to the MAD and Bias measures. The predicted values are used instead of the raw data obtained from the simulations to tease out systematic sources of the relationships from random sources.

Table 5 displays the predictions of the Farm-MAD values for the Normal distribution using the regression model showed in Table 1. Similar tables were constructed for all the assumed distributions and the three measures of accuracy. As can be seen from Table 5, the NF does not reduce the values of the Farm-MADs in any of the premium estimation procedures when yields are normally distributed. Since this result was highly consistent across all the distributions and three accuracy measures we only present tables displaying the average estimated accuracy measures across all NF for each SS and the average estimated premium across all NF and all SS (lower part of table 5) in table 6 (predicted Farm-MAD), table 7 (predicted Group-MAD) and table 8 (predicted Bias-MAD).

Regarding the magnitude of the Farm-MADs (table 6), the relative accuracy of the procedure used to estimate premiums depends on the SS. For the smallest SS of 10, which are likely to occur especially for new or small crops, M2 and M3 usually outperform (for M3 the exception is the SBN distribution) all the other methods. As the sample size increases the performance of M1 becomes comparable to M2 and M3 and in most cases better than both methods at SS=50. These results reveal the trade-off between the total number of parameters in the yield models and the accuracy of the estimated premiums with small sample sizes.

On the other extreme of performance, we have M4 which in most cases is the worst performing estimation method in terms of Farm-MAD. Even at a SS equal to 50, Farm-MADs from the estimated premiums using M4 are, on average, at least 70% off their true values. A comparison between M4 and M5 indicates that the use of the exponential tends to reduce Farm-MADs substantially.

Concerning the performance of M6 which except for the Normal distribution is the incorrect model, it is interesting to see that this method in some cases outperforms the “true” distribution models (M1, M2 and M3) or models based on historical losses (M4 and M5). This result highlights the difficulty of accurately estimating the premiums at the farm level with small samples. Of course, this result in itself is highly variable and cannot be generalized. Moreover, accuracy improvements when using the wrong model are generally lower than those obtained with the methods that use the true models or the nonparametric methods.

Table 7 presents the predicted Group-MAD values which is a measure of the variability of the average (across all farms) estimated premiums. As it would be expected the variability of the average group premiums is lower than the variability at the farm level. Across all the distributions, the SS consistently reduces Group-MADs only for M1 and M3. These two procedures are also the ones with the lowest Group-MADs at the SS=50. The relative performance of the procedures in terms of the Group-MADs is more variable at SS equal to 10 and 25, but as in the case of the Farm-MAD, M1, M2 and M3 tend to outperform other methods. On the other extreme of performance among the procedures, in terms of the Group-MAD, we have M5 which in most cases is the worst or among the worst performing methods. Group-MADs for M5 are also in every case higher than those for M4. In other words, the use of the RMA-like exponential procedure increases the variability of the group level average premiums relative to the individual historical loss premium estimation procedure. In theory, the average of the estimated premiums using the exponential obtained from the regression should be equal to the average value of the dependent variable (ER_i); however, there were several cases when the exponential was found to be higher than -1 and even positive in some of the simulation runs. In those cases, the estimated Exponential was substituted by a value of -1 which also explains the difference between the Group-MAD values obtained with M4 and M5.

Group-Bias values for the estimated premiums are shown in table 8. As explained previously, the Group-Bias measures the difference between the average (across all farms) estimated premiums and true average group premium. Similarly to Farm-MAD and Group-MAD, we only present the results of average values (across all NF) for each SS value. Consistent with the regression results, M1 and M3 are the only methods whose accuracy, as measured by the Group-Bias, is improved by the SS; hence, at the SS=50 they tend to outperform or be comparable with the best of the performing model across all distributions. However, at SS equal to 10 and 25 the relative performance of all the methods is quite variable.

There are three other aspects that can be pointed out in relation to table 8. First, the RMA-like procedure using the exponential (M5) increases the Group-Bias relative to the individual historical losses procedure (M4). Second, both M4 and M5 have in most cases a positive bias. And, thirdly the use of the wrong distribution in M6 results in group average premiums that are biased, however the magnitude and direction of the bias depends upon the underlying distribution.

In short, the analysis of the predicted accuracy measures suggests the following:

a) M1 and M3 are the only methods whose accuracy as measured by the Farm-MAD, Group-MAD and Bias-MAD is significantly and consistently improved by the SS.

b) Related with the previous results, for the largest SS (50) M1 and M3 tend to be the most accurate methods.

c) When SS equals 10 and 25 the relative performance of all the evaluated procedures is more variable, however M1, M2 and M3 are always among the most accurate procedures.

d) Consistent with the regression results, the effect of CC was more obvious in the Group-MAD values and M1, M2 and M3 where Group-MADs with CC=0.5 were about twice as big as the Group-MADs with CC=0.

e) Even though in some instances M6 outperform other procedures in terms of Group-MAD and Group-Bias, no pattern of performance of the procedure is obvious, except for the normal distribution in which case, as expected, its performance is the same as M1.

f) The use and calculation of the yield exponential (M5) to go from the group average premium to the individual premium decreases the variability of the farm level premium estimates but increases both the variability and the bias of the group premium estimates.

Conclusions and Implications

Since the Farm-MAD values measure the variability of the premium estimates at the individual (farm) level, in our opinion, they should be the focus of any analysis looking at the relative performance of insurance premium procedures. After all, a market based crop insurance program can only be successful if farmers are charged at least approximately fair premiums. Simulation results indicate that farm level premium estimates based on individual yield data and approximately correct yield models (M1, M2 and M3) are significantly more accurate than those based on historical indemnity and liability records (M4 and M5) or incorrect yield models (M6). From an implementation perspective, both types of approaches require the use of statistical models for estimation. What this result suggests is that the additional time and effort (if any) spent trying to model an appropriate yield distribution model can result in significant accuracy gains, thus reducing the potential for adverse selection problems.

Our simulation results also suggest that with small samples simpler yield models that include one mean and variance for the group (or several means and one variance) provide farm level premiums that are on average closer to the farm "true" premium than models than intend to estimate separate means and variances for each individual farm. This apparent counterintuitive result indicates than with small sample sizes (25 years or less) the added variability of estimating a large number of parameters can outweigh potential accuracy gains obtained from estimating the "correct" model which in this case includes a separate mean and variance for each farm yield model.

The simulation results also indicate that with small samples an appropriate characterization of the "group" yield distributions skewness and kurtosis might be more important than the estimation of the individual farm variances and/or means.

The RMA's current approach to farm level premium estimation can be seen as a top-down type of premium estimation approach. The procedure begins with the calculation of a county (group) premium rate. Individual farms' premiums are then calculating based on the farm's APH yields relative the group average yield. Our simulation of a similar procedure showed that in fact the accuracy of individual level farm premiums is reduced relative to a procedure where premiums are estimated based on individual indemnity and liability data (M5 versus M4). However, the simulations also showed that, contrary to expectations, this type of approach increases both the variability and bias of the group average premiums (Group-MAD and Bias-MAD).

Finally, with regard to the effect of the number of observations and number of farms on the accuracy of premium estimates, our results suggest that in general time would be better spent on trying to find more years of data than additional farms.

Table 1. Farm-MAD of Estimated Premiums: Normal Distribution

SS	NF	Correlation Coefficient 0						Correlation Coefficient 0.5					
		Premium Estimation Procedure						Premium Estimation Procedure					
		M1	**M2**	**M3**	**M4**	**M5**	**M6**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**
10	25	6.06	3.94	3.30	12.97	10.54	6.32	5.97	4.58	4.71	12.62	10.24	6.62
	50	6.39	3.89	3.37	14.92	12.38	6.81	6.17	5.67	4.37	11.48	8.18	6.49
	100	6.26	4.28	3.47	14.68	12.24	6.76	5.90	4.80	4.81	14.61	12.47	6.56
25	25	4.07	3.16	2.36	10.79	10.14	4.29	4.00	4.76	3.30	9.91	9.25	4.17
	50	4.10	4.19	3.33	12.20	11.41	4.36	4.15	4.65	3.90	8.38	7.33	4.71
	100	4.10	4.12	3.21	12.66	12.43	4.18	3.95	4.58	3.64	12.42	12.46	3.98
50	25	2.84	4.44	2.83	9.55	10.00	2.97	2.94	5.10	3.74	8.17	8.14	2.90
	50	2.91	4.17	2.96	11.44	11.73	3.06	2.84	4.37	3.57	7.16	6.72	3.13
	100	3.00	3.66	3.14	12.98	13.89	3.03	2.91	4.09	3.24	10.78	11.61	3.12

Notes: SS=sample size, NF= number of farms, CC=correlation coefficient

M1: true distribution and separate means and variances for each farm

M2: true distribution with constant mean and variance for all farms

M3: true distribution with different means but a constant variance for all farms

M4: individual farm indemnities using APH yields

M5: RMA-like procedure based on "group" indemnities and liabilities using APH yields and exponential

M6: normal distribution with separate means and variances for each farm

Table 2. Farm-MAD Regression Model Results

	M1	M2	M3	M4	M5	M6	R2	Adj-R2
Normal Distribution								
Intercept	11.518 ***	4.707 ***	4.707 ***	11.518 ***	4.707 ***	11.518 ***	0.964	0.960
ln(SS)	-2.154 ***	-0.172	-0.440 ***	-2.154 ***	-0.440 ***	-2.154 ***		
ln(NF)	-0.049	-0.049	-0.049	1.976 ***	2.108 ***	-0.049		
CC	-0.157	1.548 ***	1.548 ***	-3.831 ***	-3.831 ***	-0.157		
SBB								
Intercept	20.313 ***	5.952 ***	11.080 ***	25.203 ***	16.970 ***	12.605 ***	0.945	0.934
ln(SS)	-4.503 ***	-0.696 ***	-2.507 ***	-3.686 ***	-2.507 ***	-1.381 ***		
ln(NF)	0.144	-0.177	0.144	-0.824 ***	-0.824 ***	0.144		
CC	-0.909 **	3.317 ***	3.317 ***	-0.909 **	1.813 **	0.002		
SBA								
Intercept	12.773 ***	7.604 ***	22.775 ***	17.338 ***	12.773 ***	7.604 ***	0.728	0.706
ln(SS)	-2.338 ***	-0.978 ***	-4.559 ***	-2.338 ***	-0.978 ***	-0.978 ***		
ln(NF)	0.000	0.000	0.000	0.000	0.000	0.000		
CC	-5.422 ***	0.000	-5.422 ***	0.000	-5.422 ***	8.665 ***		
SUA								
Intercept	17.555 ***	8.499 ***	8.499 ***	22.915 ***	17.555 ***	17.555 ***	0.976	0.973
ln(SS)	-2.827 ***	-1.170 ***	-1.170 ***	-3.638 ***	-2.827 ***	-1.784 ***		
ln(NF)	-1.024 ***	-0.585 ***	-0.585 ***	0.000	-0.585 ***	0.000		
CC	1.717 ***	4.492 ***	4.492 ***	1.717 ***	5.392 ***	-0.405		
SUB								
Intercept	18.477 ***	8.071 ***	11.716 ***	19.167 ***	11.622 ***	11.662 ***	0.925	0.905
ln(SS)	-3.069 ***	-1.055 ***	-1.902 ***	-3.678 ***	-2.419 ***	-1.189 ***		
ln(NF)	-0.946 **	-0.507 *	-0.588 **	0.594 *	0.412	0.117		
CC	0.949	3.947 ***	3.822 ***	1.636 **	3.640 ***	-0.258		

Notes: SS=sample size, NF= number of farms, CC=correlation coefficient. One (*), two (**) and three (***) asterisks represent 0.10, 0.05 and 0.01 levels of statistical significance, respectively.
M1: true distribution and separate means and variances for each farm
M2: true distribution with constant mean and variance for all farms
M3: true distribution with different means but a constant variance for all farms
M4: individual farm indemnities using APH yields
M5: RMA-like procedure based on "group" indemnities and liabilities using APH yields and exponential
M6: normal distribution with separate means and variances for each farm

Table 3. Group-MAD Regression Model Results

	M1		M2		M3		M4		M5		M6		R2	Adj-R2
Normal Distribution														
Intercept	3.993	***	3.993	***	3.993	***	-2.062		-2.062		3.993	***	0.934	0.927
ln(SS)	-0.796	***	-0.299	***	-0.796	***	0.667	**	0.667	**	-0.796	***		
ln(NF)	-0.154	**	-0.154	**	-0.154	**	1.408	***	1.957	***	-0.154	**		
CC	3.036	***	1.571	***	3.036	***	5.877	***	5.877	***	3.036	***		
SBB														
Intercept	8.823	***	4.416	***	12.600	***	9.294	***	12.886	***	4.416	***	0.878	0.860
ln(SS)	-1.487	***	-0.614	***	-3.087	***	-1.487	***	-2.263	***	0.557	***		
ln(NF)	-0.588	***	-0.027		-0.027		-0.588	***	-0.588	***	-0.027			
CC	5.272	***	4.220	***	4.220	***	5.272	***	4.220	***	1.790	***		
SBA														
Intercept	13.297	***	2.336	***	25.271	***	5.518	***	13.297	***	2.336	***	0.643	0.606
ln(SS)	-2.731	***	0.000		-5.722	***	0.000		0.000		0.000			
ln(NF)	0.000		0.000		0.000		0.000		-1.662	**	0.000			
CC	-6.636	***	-0.487	*	-6.636	***	5.777	***	5.777	***	1.153	**		
SUA														
Intercept	9.540	***	9.540	***	9.540	***	9.540	***	14.844	***	9.540	***	0.901	0.893
ln(SS)	-1.368	***	-1.368	***	-1.368	***	-2.125	***	-3.212	***	-1.368	***		
ln(NF)	-0.860	***	-0.860	***	-0.860	***	0.000		0.000		-0.529	***		
CC	5.164	***	5.164	***	5.164	***	5.164	***	3.100	***	5.164	***		
SUB														
Intercept	7.785	***	6.506	***	11.407	***	2.966	***	6.506	***	2.966	***	0.873	0.854
ln(SS)	-1.141	***	-1.141	***	-2.189	***	-0.520	**	-1.141	***	0.583	***		
ln(NF)	-0.608	***	-0.364	**	-0.608	***	0.694	***	0.694	***	0.280	*		
CC	3.236	***	4.813	***	4.813	***	3.236	***	1.051	**	1.051	**		

Notes: SS=sample size, NF= number of farms, CC=correlation coefficient. One (*), two (**) and three (***) asterisks represent 0.10, 0.05 and 0.01 levels of statistical significance, respectively.

M1: true distribution and separate means and variances for each farm
M2: true distribution with constant mean and variance for all farms
M3: true distribution with different means but a constant variance for all farms
M4: individual farm indemnities using APH yields
M5: RMA-like procedure based on "group" indemnities and liabilities using APH yields and exponential
M6: normal distribution with separate means and variances for each farm

Table 4. Group-Bias Regression Model Results

	M1	M2	M3	M4	M5	M6	R2	Adj-R2
Normal Distribution								
Intercept	-1.714 **	0.942	-2.304 ***	-1.714 **	0.846	0.846	0.946	0.940
ln(SS)	0.452 ***	0.452 ***	0.452 ***	0.452 ***	0.452 ***	-0.107		
ln(NF)	-0.026	-0.026	-0.026	2.134 ***	2.134 ***	-0.026		
CC	0.165	0.165	0.165	-5.008 ***	-5.008 ***	0.165		
SBB								
Intercept	-11.778 ***	-2.394 ***	7.360 ***	1.098 ***	8.611 ***	-1.681 *	0.879	0.865
ln(SS)	2.207 ***	0.000	-2.893 ***	0.000	-1.806 ***	-1.321 ***		
ln(NF)	0.920 ***	0.000	0.920 ***	0.000	0.000	0.000		
CC	-0.856 **	-0.856 **	-0.856 **	-0.856 **	-0.856 **	-0.856 **		
SBA								
Intercept	7.925 **	0.883	19.023 ***	19.023 ***	19.023 ***	0.883	0.559	0.514
ln(SS)	-1.526 **	0.000	-4.294 ***	0.000	0.000	0.000		
ln(NF)	0.370	0.370	0.370	-4.083 ***	-3.516 ***	0.370		
CC	-9.491 ***	-0.641 *	-9.491 ***	9.242 ***	9.242 ***	-0.641 *		
SUA								
Intercept	3.380 ***	-0.280	3.380 ***	-0.280	7.142 ***	-0.280	0.768	0.739
ln(SS)	-0.017	-0.017	-0.436 **	-0.436 **	-1.868 ***	-0.017		
ln(NF)	-0.809 ***	-0.252 **	-0.252 **	0.634 ***	0.634 ***	0.034		
CC	-1.157 **	-0.302	-0.302	-1.157 **	-2.709 **	-1.157 **		
SUB								
Intercept	4.137 **	1.108	6.502 ***	-3.811 ***	1.108	-3.811 ***	0.931	0.922
ln(SS)	-0.448 ***	-0.448 ***	-1.684 ***	0.607 **	-0.448 ***	-0.629 ***		
ln(NF)	-0.554 *	-0.023	-0.023	1.428 ***	1.428 ***	-0.023		
CC	-2.299 ***	-2.299 ***	0.263	-2.299 ***	-2.299 ***	0.263		

Notes: SS=sample size, NF= number of farms, CC=correlation coefficient. One (*), two (**) and three (***) asterisks represent 0.10, 0.05 and 0.01 levels of statistical significance, respectively.

M1: true distribution and separate means and variances for each farm

M2: true distribution with constant mean and variance for all farms

M3: true distribution with different means but a constant variance for all farms

M4: individual farm indemnities using APH yields

M5: RMA-like procedure based on “group” indemnities and liabilities using APH yields and exponential

M6: normal distribution with separate means and variances for each farm

Table 5. Predicted Farm-level Mean Absolute Differences of Estimated Premiums: Normal Distribution

SS	NF	Correlation Coefficient											
		0						0.5					
		Premium Estimation Procedure						Premium Estimation Procedure					
		M1	**M2**	**M3**	**M4**	**M5**	**M6**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**
10	25	5.88	4.21	3.36	10.73	7.92	5.83	5.85	4.99	4.14	12.85	10.05	5.79
	50	6.36	4.12	3.50	14.29	11.94	6.36	6.29	4.89	4.28	12.37	10.03	6.29
	100	6.33	4.09	3.47	15.66	13.40	6.33	6.25	4.86	4.24	13.74	11.49	6.25
25	25	4.43	4.00	3.13	10.94	10.08	4.43	4.35	4.77	3.91	9.03	8.16	4.35
	50	4.39	3.96	3.10	12.31	11.54	4.39	4.31	4.74	3.87	10.40	9.62	4.31
	100	4.36	3.93	3.06	13.68	13.00	4.36	4.28	4.70	3.84	11.77	11.08	4.28
50	25	2.93	3.88	2.83	9.45	9.77	2.93	2.85	4.65	3.60	7.54	7.86	2.85
	50	2.90	3.84	2.79	10.82	11.23	2.90	2.82	4.62	3.57	8.91	9.32	2.82
	100	2.86	3.81	2.76	12.19	12.69	2.86	2.79	4.58	3.53	10.28	10.78	2.79
All	**25**	4.59	4.01	3.17	11.10	10.11	4.59	4.51	4.78	3.94	9.19	8.19	4.51
	50	4.55	3.97	3.13	12.47	11.57	4.55	4.47	4.75	3.91	10.56	9.66	4.47
	100	4.52	3.94	3.10	13.84	13.03	4.52	4.44	4.71	3.87	11.93	11.12	4.44
10	**All**	6.36	4.12	3.50	14.29	11.94	6.36	6.29	4.89	4.28	12.37	10.03	6.29
25		4.39	3.96	3.10	12.31	11.54	4.39	4.31	4.74	3.87	10.40	9.62	4.31
50		2.90	3.84	2.79	10.82	11.23	2.90	2.82	4.62	3.57	8.91	9.32	2.82
All	**All**	4.55	3.98	3.13	12.47	11.57	4.55	4.47	4.75	3.91	10.56	9.66	4.47

Notes: SS=sample size, NF= number of farms.

M1: true distribution and separate means and variances for each farm

M2: true distribution with constant mean and variance for all farms

M3: true distribution with different means but a constant variance for all farms

M4: individual farm indemnities using APH yields

M5: RMA-like procedure based on "group" indemnities and liabilities using APH yields and exponential

M6: normal distribution with separate means and variances for each farm

Table 6. Predicted Farm-level Mean Absolute Differences of Estimated Premiums

Distribution	SS	NF	Correlation Coefficient											
			0						0.5					
			Premium Estimation Procedure						Premium Estimation Procedure					
			M1	M2	M3	M4	M5	M6	M1	M2	M3	M4	M5	M6
Normal	All	10	6.36	4.12	3.50	14.29	11.94	6.36	6.29	4.89	4.28	12.37	10.03	6.29
		25	4.39	3.96	3.10	12.31	11.54	4.39	4.31	4.74	3.87	10.40	9.62	4.31
		50	2.90	3.84	2.79	10.82	11.23	2.90	2.82	4.62	3.57	8.91	9.32	2.82
	All	All	4.55	3.98	3.13	12.47	11.57	4.55	4.47	4.75	3.91	10.56	9.66	4.47
SBB	All	10	10.51	3.66	5.87	13.49	7.97	9.99	10.05	5.32	7.53	13.04	8.88	9.99
		25	6.38	3.02	3.58	10.11	5.68	8.73	5.93	4.68	5.23	9.66	6.58	8.73
		50	3.26	2.54	1.84	7.56	3.94	7.77	2.81	4.20	3.50	7.10	4.85	7.77
	All	All	6.72	3.07	3.76	10.39	5.86	8.83	6.26	4.73	5.42	9.93	6.77	8.83
SBA	All	10	7.39	5.35	12.28	11.95	10.52	5.35	4.68	5.35	9.57	11.95	7.81	9.68
		25	5.25	4.45	8.10	9.81	9.62	4.45	2.53	4.45	5.39	9.81	6.91	8.79
		50	3.63	3.78	4.94	8.19	8.95	3.78	0.91	3.78	2.23	8.19	6.23	8.11
	All	All	5.42	4.53	8.44	9.98	9.70	4.53	2.71	4.53	5.73	9.98	6.98	8.86
SUA	All	10	7.04	3.52	3.52	14.54	8.76	13.45	7.90	5.76	5.76	15.40	11.45	13.25
		25	4.45	2.44	2.44	11.20	6.17	11.81	5.31	4.69	4.69	12.06	8.86	11.61
		50	2.49	1.63	1.63	8.68	4.21	10.58	3.35	3.88	3.88	9.54	6.90	10.37
	All	All	4.66	2.53	2.53	11.47	6.38	11.95	5.52	4.78	4.78	12.33	9.07	11.74
SUB	All	10	7.71	3.66	5.04	13.02	7.67	9.38	8.18	5.63	6.95	13.84	9.49	9.25
		25	4.90	2.69	3.29	9.65	5.45	8.29	5.37	4.66	5.21	10.47	7.27	8.16
		50	2.77	1.96	1.98	7.10	3.77	7.47	3.24	3.93	3.89	7.92	5.59	7.34
	All	All	5.12	2.77	3.44	9.92	5.63	8.38	5.60	4.74	5.35	10.74	7.45	8.25

Notes: SS=sample size, NF= number of farms
M1: true distribution and separate means and variances for each farm
M2: true distribution with constant mean and variance for all farms
M3: true distribution with different means but a constant variance for all farms
M4: individual farm indemnities using APH yields
M5: RMA-like procedure based on "group" indemnities and liabilities using APH yields and exponential
M6: normal distribution with separate means and variances for each farm

Table 7. Predicted Group-level Mean Absolute Differences of Estimated Premiums

Distribution	SS	NF	Correlation Coefficient 0: Premium Estimation Procedure M1	M2	M3	M4	M5	M6	Correlation Coefficient 0.5: Premium Estimation Procedure M1	M2	M3	M4	M5	M6
Normal	All	10	1.56	2.70	1.56	4.98	7.13	1.56	3.08	3.49	3.08	7.92	10.07	3.08
		25	0.83	2.43	0.83	5.59	7.74	0.83	2.35	3.21	2.35	8.53	10.68	2.35
		50	0.28	2.22	0.28	6.05	8.20	0.28	1.80	3.01	1.80	8.99	11.14	1.80
	All	All	0.89	2.45	0.89	5.54	7.69	0.89	2.41	3.24	2.41	8.48	10.63	2.41
SBB	All	10	3.10	2.90	5.39	3.57	5.37	5.59	5.73	5.01	7.50	6.20	7.48	6.49
		25	1.74	2.33	2.56	2.21	3.30	6.10	4.37	4.44	4.67	4.84	5.41	7.00
		50	0.70	1.91	0.42	1.17	1.73	6.49	3.34	4.02	2.53	3.81	3.84	7.38
	All	All	1.85	2.38	2.79	2.32	3.47	6.06	4.48	4.49	4.90	4.95	5.58	6.95
SBN	All	10	7.01	2.34	12.10	5.52	6.79	2.34	3.69	2.09	8.78	8.41	9.68	2.91
		25	4.51	2.34	6.85	5.52	6.79	2.34	1.19	2.09	3.54	8.41	9.68	2.91
		50	2.61	2.34	2.89	5.52	6.79	2.34	-0.70	2.09	-0.43	8.41	9.68	2.91
	All	All	4.71	2.34	7.28	5.52	6.79	2.34	1.39	2.09	3.96	8.41	9.68	2.91
SUI	All	10	3.03	3.03	3.03	4.65	7.45	4.32	5.61	5.61	5.61	7.23	9.00	6.90
		25	1.77	1.77	1.77	2.70	4.50	3.07	4.35	4.35	4.35	5.28	6.05	5.65
		50	0.82	0.82	0.82	1.23	2.28	2.12	3.41	3.41	3.41	3.81	3.83	4.70
	All	All	1.87	1.87	1.87	2.86	4.74	3.17	4.46	4.46	4.46	5.44	6.29	5.75
SUR	All	10	2.78	2.46	3.99	4.48	6.59	5.41	4.40	4.86	6.39	6.10	7.12	5.93
		25	1.73	1.41	1.98	4.01	5.55	5.94	3.35	3.82	4.39	5.63	6.07	6.47
		50	0.94	0.62	0.46	3.65	4.76	6.34	2.56	3.03	2.87	5.27	5.28	6.87
	All	All	1.82	1.50	2.14	4.05	5.63	5.90	3.44	3.90	4.55	5.67	6.16	6.42

Notes: SS=sample size, NF= number of farms

M1: true distribution and separate means and variances for each farm

M2: true distribution with constant mean and variance for all farms

M3: true distribution with different means but a constant variance for all farms

M4: individual farm indemnities using APH yields

M5: RMA-like procedure based on "group" indemnities and liabilities using APH yields and exponential

M6: normal distribution with separate means and variances for each farm

Table 8. Predicted Group-level Bias of Estimated Premiums

Distribution	SS	NF	Correlation Coefficient											
			0						0.5					
			Premium Estimation Procedure						Premium Estimation Procedure					
			M1	M2	M3	M4	M5	M6	M1	M2	M3	M4	M5	M6
Normal	All	10	-1.08	1.60	-1.66	4.35	6.81	0.62	-1.03	1.65	-1.61	7.37	9.83	0.68
		25	-0.35	2.33	-0.93	5.08	7.54	0.32	-0.30	2.39	-0.87	8.11	10.57	0.37
		50	0.20	2.89	-0.37	5.64	8.10	0.09	0.26	2.94	-0.32	8.66	11.12	0.14
	All	All	-0.41	2.27	-0.99	5.02	7.48	0.34	-0.36	2.33	-0.93	8.05	10.51	0.40
SBB	All	10	-2.70	-2.39	4.04	1.10	3.82	-4.86	-3.13	-2.82	3.61	0.67	3.39	-5.29
		25	-1.50	-2.39	2.33	1.10	3.02	-5.66	-1.92	-2.82	1.90	0.67	2.59	-6.09
		50	0.51	-2.39	-0.52	1.10	1.69	-6.99	0.08	-2.82	-0.95	0.67	1.26	-7.42
	All	All	-1.23	-2.39	1.95	1.10	2.84	-5.84	-1.66	-2.82	1.52	0.67	2.41	-6.27
SBA	All	10	5.86	2.33	10.58	3.05	5.27	2.33	1.11	2.01	5.84	7.67	9.89	2.01
		25	4.46	2.33	6.65	3.05	5.27	2.33	-0.29	2.01	1.90	7.67	9.89	2.01
		50	3.40	2.33	3.67	3.05	5.27	2.33	-1.34	2.01	-1.07	7.67	9.89	2.01
	All	All	4.57	2.33	6.97	3.05	5.27	2.33	-0.17	2.01	2.22	7.67	9.89	2.01
SUA	All	10	0.18	-1.30	1.39	1.19	5.32	-0.19	-0.40	-1.46	1.24	0.62	3.96	-0.76
		25	0.16	-1.32	0.99	0.79	3.61	-0.20	-0.42	-1.47	0.84	0.22	2.25	-0.78
		50	0.15	-1.33	0.69	0.49	2.31	-0.21	-0.43	-1.48	0.54	-0.09	0.96	-0.79
	All	All	0.16	-1.32	1.02	0.82	3.75	-0.20	-0.42	-1.47	0.87	0.25	2.39	-0.78
SUB	All	10	0.94	-0.01	2.54	3.17	5.66	-5.35	-0.21	-1.16	2.67	2.02	4.51	-5.22
		25	0.53	-0.42	0.99	3.73	5.25	-5.93	-0.62	-1.57	1.12	2.58	4.10	-5.80
		50	0.22	-0.74	-0.17	4.15	4.94	-6.36	-0.93	-1.88	-0.04	3.00	3.79	-6.23
	All	All	0.56	-0.39	1.12	3.68	5.28	-5.88	-0.59	-1.54	1.25	2.53	4.13	-5.75

Notes: SS=sample size, NF= number of farms

M1: true distribution and separate means and variances for each farm

M2: true distribution with constant mean and variance for all farms

M3: true distribution with different means but a constant variance for all farms

M4: individual farm indemnities using APH yields

M5: RMA-like procedure based on "group" indemnities and liabilities using APH yields and exponential

M6: normal distribution with separate means and variances for each farm

References

Babcock, B.A. 2008. "Corn Belt Contributions to the Crop Insurance Industry," *Iowa Ag Review* 14(2):1-3, 11.

Deng, X. H., B. J. Barnett, and D. V. Vedenov. 2007. "Is there a Viable Market for Area-based Crop Insurance?" *American Journal of Agricultural Economics* 89(2):508-519.

Goodwin, B.K. "Premium Rate Determination in the Federal Crop Insurance Program: What Do Averages Have to Say About Risk?" *Journal of Agricultural and Resource Economics.* Vol. 19, No. 2: p. 382-395.

Goodwin, B.K. and L. Vado. "Public Responses to Agricultural Disasters: Rethinking the Role of Government." *Canadian Journal of Agricultural Economics* 55(4):399-417.

Glauber, J. W. 2004. "Crop Insurance Reconsidered." *American Journal of Agricultural Economics* 86(5):1179–1195.

Harri, A., K.H. Coble, C. Erdem, and T.O. Knight. 2005. "Crop Yield Normality: A Reconciliation of Previous Research." Working Paper, Department of Agricultural Economics, Mississippi State University, Starkville, Mississippi.

Ker, A.P. and K. Coble. 2003. "Modeling Conditional Yield Densities." *American Journal of Agricultural Economics* 85(2):291-304.

Milliman and Robertson, Inc. 2000. Actuarial Documentation of Multiple Peril Crop Insurance Ratemaking Procedures. Consulting report prepared for the Risk Management Agency, USDA. Kansas City, MO: USDA-Risk Management Agency.

Nelson, C.H. and P.V. Preckel. 1989. "The Conditional Beta Distribution as a Stochastic Production Function." *American Journal of Agricultural Economics* 71(2):370-378.

Ramirez, O.A. and T. McDonald. “An Empirically-Grounded Comparison of the Johnson System versus the Beta as Crop Yield Distribution Models.” Selected poster presented at the annual meeting of the American Agricultural Economics Association, Portland, Oregon, July 29-31 2007. Paper posted in AgEcon Search (http://agecon.lib.umn.edu/).

Ramirez, O.A. and T. McDonald. “The Expanded and Re-parameterized Johnson System: A Most Flexible Crop-Yield Distribution Model.” Paper presented at the annual meeting of the American Agricultural Economics Association, Long Beach, California, July 23-26 2006 and posted in AgEcon Search (http://agecon.lib.umn.edu/).

Ramirez, O.A. 1997. “Estimation and Use of a Multivariate Parametric Model for Simulating Heteroskedastic, Correlated, Non-Normal Random Variables: The Case of Corn-Belt Corn, Soybeans and Wheat Yields.” *American Journal of Agricultural Economics* 79(1):191-205.

Skees, J.R. and M.R. Reed. 1986. “Rate-Making and Farm-Level Crop Insurance: Implications for Adverse Selection.” *American Journal of Agricultural Economics* 68:653-659.

Stute, W. 1999. “Nonlinear Censored Regression.” Statistica Sinica 9:1089-1102.

Taylor, C.R. 1990. “Two Practical Procedures for Estimating Multivariate Non-Normal Probability Density Functions.” *American Journal of Agricultural Economics* 72(1):210-217.

Worth, T. 2008. “Discussion: The 2007 Farm Bill and Crop Insurance: Implications for Crop Producers in the South.” *Journal of Agricultural and Applied Economics* 40(2): p. 461-464.