Can Crop Insurance Premiums Be Reliably Estimated?

Octavio A. Ramirez, Carlos E. Carpio, and Roderick M. Rejesus

This paper develops and applies a methodology to assess the accuracy of historical loss-cost rating procedures, similar to those used by the U.S. Department of Agriculture's Risk Management Agency (RMA), versus alternative parametric premium estimation methods. It finds that the accuracy of loss-cost procedures leaves much to be desired, but can be markedly improved through the use of alternative methods and increased farm-level yield sample sizes. Evidence suggests that the high degree of inaccuracy in crop insurance premium estimations through historical loss-cost procedures identified in the paper might be a major factor behind the need for substantial government subsidies to keep the program solvent.

Key Words: agricultural subsidies, crop insurance premium estimation, loss-cost procedures, Risk Management Agency

The U.S. crop insurance program is a joint effort of the federal government and private insurance companies that sell policies to farmers backed by reinsurance provided by the Federal Crop Insurance Corporation. The Risk Management Agency (RMA), a division of the U.S. Department of Agriculture, administers this insurance program. The traditional product offered by the RMA, which is the focus of this paper, is a farm-level, multiple-peril, crop yield insurance policy [called the MPCI or Actual Production History (APH) policy]. This policy protects against low yield and crop quality losses due to adverse weather and unavoidable damage from insects and disease (Barnett 2000).

During the past 15 years, the federal government has increasingly looked to crop insurance as a possible alternative to the historical "disaster relief" payments that are made to farmers when crop yields are drastically reduced due to widespread bad weather, pest outbreaks, or other adverse events. Therefore, through the RMA, it has tried to promote participation by subsidizing the premiums paid by farmers. In 2009, the U.S. crop insurance program covered close to 265 million acres, assuming nearly $80 billion in liabilities. This breadth of coverage has been obtained through increased subsidies over time, with producers as a whole now paying only about 40 percent of the total premiums required to keep the program solvent.[1] The need for such large subsidies to achieve high levels of producer participation has for the most part been attributed to "adverse selection" (Harwood et al. 1999).

Specifically, it has been hypothesized that farmers are better able to ascertain what their actuarially fair premiums are than the RMA, and they tend to participate only if they feel that it is to their economic advantage. As a result, the program is loaded with producers whose fair premiums are lower than what they are being charged. In short, the root of the adverse selection problem is the RMA's inability to precisely estimate the actuarially fair premiums that should be charged to individual producers. In addition to their impact on the actuarial performance of the crop insurance program, incorrect rates can affect the producers' economic welfare and the incentives and returns to the private companies that sell federal crop insurance at those rates.

Octavio A. Ramirez is Professor and Head of the Department of Agricultural and Applied Economics at the University of Georgia, Athens, Georgia. Carlos E. Carpio is Assistant Professor in the Department of Applied Economics and Statistics at Clemson University, Clemson, South Carolina. Roderick M. Rejesus is Associate Professor in the Department of Agricultural and Resource Economics at North Carolina State University, Raleigh, North Carolina.

[1] See http://www.rma.usda.gov/data/sob.html.

Although an extensive and highly relevant body of work on crop insurance program rating has been published in the agricultural economics literature to date, the question of how accurately crop insurance premiums can be estimated through historical loss-cost procedures (i.e., the basic approach used by RMA) and other proposed methods remains largely unanswered. Specifically, no study has quantified the magnitude of the inaccuracy in the premium estimates obtained under these alternative rating methods. The reason for this gap in the literature might be that, in principle, the "true" (i.e., actuarially correct) premium corresponding to any particular farmer is unknown, making an actual comparison between estimated and true rates unfeasible.

This research makes such a comparison possible through simulation methods. Specifically, the analysis is based on yield distributions previously estimated on the basis of one of the most comprehensive farm-level datasets in the United States (Sherrick et al. 2004) and recently developed parametric modeling procedures. Since these procedures are flexible enough to accommodate a wide variety of distributional shapes (Ramirez, McDonald, and Carpio 2010, Ramirez and McDonald 2006a, Ramirez, Misra, and Field 2003), the estimated distributions should sufficiently resemble the true underlying yield densities to make the analyses realistic. The estimated distributions are then assumed to be the true data-generating processes and used to simulate yield datasets for the analyses. With this construct, the true premiums can be computed with near certainty on the basis of large simulated datasets. In addition, premium estimates can be obtained under a variety of methods and small sample sizes drawn from those same distributions, making the desired comparisons possible.

Through such procedures, this paper assesses the accuracy of various rating methods by comparing the estimated rates to true rates under different underlying yield distributions and common data availability scenarios and conditions such as sample size, the number of farms from which data is available, and the level of yield correlation across farms. The rating methods evaluated include the historical loss-cost procedures that rely on liability and indemnity data (similar to those currently used by the RMA) and alternative parametric methods that use estimated yield distribu-

tion models to simulate the expected losses. The resulting statistics are used to preliminarily assess the impact of premium estimation inaccuracy on producer participation rates, the relative levels of premium subsidy needed to achieve those rates, and the resulting increases in government costs for the particular case of Illinois corn farmers.

In addition to advancing a methodology to quantify the current levels of crop insurance premium estimation inaccuracy, this article explores the potential improvements in precision that could be expected from applying alternative estimation procedures and using larger yield sample sizes. It also provides preliminary evidence to suggest that premium estimation inaccuracy could, by itself, be the cause of the high loss ratios and resulting need for elevated government subsidies since the program's inception.

Theoretical Framework

A farmer participating in the APH program selects one of several possible yield guarantees (α) and some price guarantee level p_g. The expected value of indemnity I for coverage at the $\alpha \times 100$ percent of the mean (M) farm-level yield is given by

$$E_y[I] = \int_0^{\alpha M} p_g (\alpha M - y) f(y) dy \,,$$

where E_y is the expectations operator and $f(y)$ the probability density function of yields. Knowledge of $E_y[I]$ is important for both the farmer and the insurer as they make their decisions to buy or sell a yield insurance product. For example, a risk-neutral farmer will purchase insurance only if $E_y[I]$ is higher than the premium charged (Coble et al. 2009). From the insurer's perspective, $E_y[I]$ is the actuarially fair premium, i.e., the one it needs to charge to avoid an expected loss. Since $E_y[I]$ is unknown, it has to be estimated by both farmers and insurers and, therefore, is subject to sampling variability. Additionally, given that the type, quality, and quantity of information available to these two parties are markedly different, the amounts of variability in their estimated $E_y[I]$ are also likely to differ. For the remainder of the paper, since the analysis is conducted from the insurer's perspective, the actuarially fair premiums ($E_y[I]$) are also referred to as the "true" premiums.

Methods and Procedures

The procedures utilized include the following major steps: (i) selection of a representative set of five estimated yield distributions, which are then assumed to be the true densities, (ii) simulation of large datasets from each of these five yield distributions, (iii) calculation of the true premiums based on those datasets, (iv) estimation of premiums using small samples drawn from the same distributions, and (v) comparison of true versus estimated premiums.

Selection of Yield Distributions

The yield distributions used for this research were selected from those estimated by Ramirez, McDonald, and Carpio (2010) using data from the University of Illinois Endowment Farms (Sherrick et al. 2004). The dataset contains yields from 26 representative corn farms located in twelve counties across that state. Ramirez, McDonald, and Carpio (2010) use this data to estimate models for those 26 yield distributions that are as realistic as possible. To this end, they utilize 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 (Ramirez, McDonald, and Carpio 2010, Ramirez and McDonald 2006a, Ramirez, Misra, and Field 2003). This property makes the Johnson system preferable for use in this research to other less flexible distributions such as the Beta or Gamma which allow for only very limited MVSK combinations (Ramirez and McDonald 2006a). Another advantage of using Ramirez, McDonald, and Carpio (2010) results is that they identify a variety of distributional shapes that span over a substantial area of the theoretically feasible skewness-kurtosis (SK) space.[2] A thoughtfully selected subset of these 26 models should, therefore, be represen-

tative of the breadth of distributional shapes that could be encountered in practice.

In their analyses, Ramirez, McDonald, and Carpio (2010) estimate normal, S_U, and S_B models for each of the 26 yield series, which include quadratic and linear time trends for the means and standard deviations respectively. They then conduct likelihood ratio tests which reject the normality hypothesis in 20 of the 26 cases ($\alpha = 0.10$). All non-normal distributions are found to be left-skewed, which is consistent with previous literature on corn yields (Nelson and Preckel 1989, Taylor 1990, Ramirez 1997, Ker and Coble 2003, Harri et al. 2005). Out of the 20 cases that are classified as non-normal, the S_B models exhibit the highest maximum log-likelihood function value in 14 cases and the S_U models in six cases.

Five of the distributions estimated by Ramirez, McDonald, and Carpio (2010) are chosen for the purposes of this research. These include one normal, two S_U's, and two S_B's. The S_U's and S_B's are selected to have (i) low skewness (-0.833) and high positive kurtosis (396.654) (S_{UA}), (ii) low skewness (-0.09) and high negative kurtosis (-1.21) (S_{BA}), (iii) moderately negative skewness (-2.10) and moderately positive kurtosis (10.01) (S_{UB}), and (iv) negative skewness (-2.77) and positive kurtosis (14.25) (S_{BB}). That is, they are representative of the SK spectrum of the 26 distributional shapes identified by Ramirez, McDonald, and Carpio (2010) to be associated with empirical farm-level yield data. For the purposes of this research, these are assumed to be the true distributional shapes underlying five typical yield data-generating processes.

Simulation of Yields

The next step in the analysis requires simulation of data from the five selected distributions (normal, S_{UA}, S_{UB}, S_{BA}, S_{BB}). The simulation formulas (Ramirez and McDonald 2006b) are

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

for the S_U,

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

---

[2] There are other flexible non-parametric and semi-parametric approaches that could be used as well (Ker and Goodwin 2000, Ker and Coble 2003). However, given that the parametric methods utilized are easily tractable, perform relatively well in small samples, and are flexible enough to approximate a wide variety of distributional shapes (Lu et al. 2008), they are believed to suffice for the purposes of this study.

for the S_B, and

(3) $$S_N = \sigma Z + M$$

for the normal distribution. 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 (available from the authors upon request). The skewness and kurtosis parameters estimated by Ramirez, McDonald, and Carpio (2010) 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 ("number of farms") sets of mean and standard deviation parameters are assumed to be drawn from uniform distributions with ranges of 140 to 180 bushels per acre and 25 to 35 bushels per acre, respectively. These are consistent with the range of means and variances estimated by those authors for their 26 estimated corn yield distributions projected to year 2010 using their estimated mean and variance trend equations.

The reason for this framework is to explore a hypothetical situation where one observes yields from a number of farms (NF) within the same county (or other aggregate rating unit), which have different means 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 analyses.

Calculation of True Premium Values

The near-exact actuarially fair crop insurance premiums corresponding to each of the five assumed yield distributions are then computed for the typical 65 percent APH coverage level on the basis of large datasets (100,000 draws) simulated from those distributions (note that an infinite sample size would be needed to compute the exact fair premium). 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 $0.65 \times \bar{Y}$, the difference $(0.65 \times \bar{Y} - Y_i)$ is multiplied by the assumed price guarantee ($2.2/bushel in this case). Otherwise the observation is discarded.

Then the 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. Since these are the actuarially fair premiums corresponding to the assumed distributions, for the purposes of this research they are considered to be the true premiums (i.e., $E_y[I]$). Given that 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." Runs for three different NF values (100, 50, and 25) are conducted. Therefore, the final output is 100, 50, and 25 sets of true premiums for each of the distributions in the analysis.

Premium Estimation

The next step is to estimate premium rates under realistic field conditions. To this effect, small samples of size 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 originally 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 following the general procedure outlined by Ramirez (1997) for the case of the S_U family. A CC of zero is a natural choice for the limit scenario where yields are independently distributed across farms. A CC of 0.5 is consistent with levels of correlation estimated in previous literature (Ramirez, Misra, and Field 2003).

The unit of analysis is a particular NF-SS-CC combination. Therefore, for each distribution there are $3 \times 3 \times 2 = 18$ units of analysis. An ideal situation for estimating farm-level rates might 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 considered in the paper is to have data on 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 final step is to use the parametric yield distribution and the loss-cost approaches to estimate the actuarially fair premiums under all of these scenarios.

The parametric yield distribution estimation approach. In this approach, the previously discussed Johnson system and the data corresponding to each unit of analysis are used to estimate the joint yield distributions by maximum likelihood (ML) estimation procedures (Ramirez 1997, Ramirez, McDonald, and Carpio 2010). Three alternative joint probability distribution models are specified and estimated: one with separate means and variances for each farm (M1), one with the same mean and variance for all farms (M2), and one with different means but the same variance (M3). Note that M1 is the theoretically correct model, i.e., it mimics the data-generating process. Based on intuition, it is hypothesized that the increased parsimony in M3 and/or M2 (relative to M1) might help improve premium estimation accuracy when working with small sample sizes. As in the data-generating process, the skewness and kurtosis are assumed constant across farms.

When needed, programming measures were taken to alleviate convergence problems, such as placing a linear constraint on the direction taken at each iteration, switching between two algorithms depending on progress towards convergence, automatically altering the grid-search to a new direction if all methods failed to compute a direction for the next step, increasing the radius of the grid-search, using five instead of two points for computing the numerical derivatives, properly scaling so that the diagonal elements of the Hessian matrix were roughly equal, running trials to determine the best starting parameter values for the final runs, and placing bounds to the parameter estimates (making sure that convergence did not occur at a boundary). After these measures were selectively taken for each particular distribution-NF-SS-CC-method scenario, convergence problems were minimal. Then, following the same procedure used to compute the true premiums, the resulting models were used to jointly simulate yield draws ($n = 100,000$) and calculate the actuarially fair premium corresponding to each of the farms in the unit of analysis. The process is repeated for all 18 units of analysis associated with each of the five distributional shapes under consideration.

As an example of the application of the parametric approach, consider the scenario where the assumed distribution is S_{UA}, NF = 25, and CC = 0.50. First, the S_{UA} skewness and kurtosis parameters estimated on the basis of the original field data plus the NF (randomly) assumed mean and variance parameters plus the assumed CC = 0.50 value are used to jointly simulate NF large (100,000 observation) samples with correlation coefficient CC = 0.50, which are used to compute the "true" premiums for those 25 "farms" using the previously described procedure.

Next, those same parameters are used to simulate small samples from the same joint yield distribution. For SS = 10, for example, a sample of 10 observations is generated for each of the NF = 25 farms that share the same (S_{UA}) skewness and kurtosis parameters and have moderately correlated yields (CC = 0.50). That sample (250 observations) is used to estimate the parameters of a joint S_U pdf. In the case of M1, for example, this pdf would have 25 mean parameters (one per farm), 25 variance parameters (one per farm), two non-normality parameters (same for all farms), and a correlation coefficient "connecting" the 25 marginal densities. These are estimates for the true parameters "assumed" in the initial step and, therefore, the resulting pdf is an estimate for what has been assumed to be the true underlying yield distribution.

The 53 parameter estimates (i.e., the estimated pdf) is then used to simulate $100,000 \times 25$ correlated yield draws, which as in the case of the true premium are needed to compute the estimated premiums for each of those NF = 25 farms. To summarize, for the parametric methods, the difference between the computation of the true versus the estimated premiums is that in the case of the true premiums the pdf used to simulate the $n = 100,000$ draws is based on the "true" parameter values, while when estimating the premiums the pdf used for the simulation contains the parameter estimates obtained from the M1, M2, or M3 models. Also note that in the case of the non-parametric methods—M4 and M5—no pdf is estimated. Instead, the RMA protocols discussed in the following section are applied to the SS observations from each of the NF farms in the sample (the RMA methods make no direct allowances for the cross-farm correlation) to compute the premium estimates. Thus, the premium estimates in all 5 methods are based on the same SS × NF simulated yield observations.

The non-parametric historical loss-cost approach. As previously explained, some of the data used to estimate the yield distributions under

the parametric approach (above) is utilized to compute farm-level premiums through two different historical loss-cost procedures. The first method (M4) is based solely on the individual farm's yield data. The premium for coverage at the 0.65 × 100 percent of the APH level is given by

$$(4) \qquad ER_i = [\sum_{t=1}^{SS} P_g Y_{it}^* / SS],$$

where

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

where ER_i is farm i's empirical APH-based premium, P_g is the guaranteed price (which is assumed constant), SS is the sample size, t denotes the year, Y_{it} is the observed yield for farm i in year t, and APH_{it} is an estimate of farm i's mean yield (μ_i) in year t. Equation (4) is similar to the empirical rate presented in Skees and Reed (1986) and Goodwin (1994). However, the mean yield in their equations (μ_i), which is unknown, is replaced by APH_{it}. Our procedure to calculate APH_{it} follows the method used by the RMA. At the beginning of the historical period, when a farmer enters the program, the RMA assigns a transitional yield (t-yield) based on the county average. That is, the RMA APH yields are not entirely based on the observed farm-level yields during the first four years of history. For our analysis, APH_{it} yields were simulated as follows: APH_{i1} for all i's was the average yield of a different batch of yield simulations, which is meant to simulate the county average during previous years (t-yield). APH_{i2} was the first simulated yield value plus three times the t-yield, divided by four. APH_{i3} and APH_{i4} are analogously calculated. Thereafter, APH_{it} (j=5,…,n) is computed using the average of all simulated yield values only.

The rate calculation based on farm-level yields only (M4) is included for two reasons. First, because it is needed to compute the farm-level losses required to simulate historical county indemnities for the second loss-cost premium estimation procedure being evaluated (M5). Also, from a statistical perspective, M4 is a non-parametric procedure that uses APH yields instead of the average yields for the entire sample. Therefore, it is of interest to compare M4 with the parametric proce-

dures (M1, M2, and M3). The second non-parametric method (M5) also incorporates historical aggregate rating unit loss and liability information using the main equation underlying the RMA rate-making procedure (Milliman and Robertson, Inc. 2000):

$$(5) \qquad G\_PR_i = P_g \times 0.65 \times APH_{iSS} \times CPR \times (\frac{APH_{iSS}}{Yavc})^{\text{Exp}},$$

where G_PR_i is the farm i county-based premium rate, P_g is the guaranteed price, CPR is the county rate, Exp (the *Exponential*) is an exponent whose value is usually less than -1, and APH_{iSS} and $Yavc$ are the APH yield for farm i and county average yield, respectively (Milliman and Robertson, Inc. 2000). Also note that in the RMA literature, the ratio

$$\frac{APH_{iSS}}{Yavc}$$

is usually called the "*yield ratio*" and that both APH_{iSS} and $Yavc$ are calculated using the entire sample of simulated yields (SS).

Although this is a simplified version of the equation used by the RMA, it includes all the elements that are central for our analysis. The logic underlying equation (5) is that the individual farm-level premiums can be established using the county rate (CPR) as the baseline. The *Exponential* (*Exp*) is used so that farmers with yields that are above the area's average pay lower premiums and vice versa (Knight 2000).[3] The calculation of CPR is based on the simulated farm-level indemnities and liabilities (i subscript) for each time period (t subscript). In year t, for example, the simulated indemnity, liability, and CPR for the NF group of farms are

$$(6) \qquad \text{Indemnity}_t = \sum_{i=1}^{NF} P_g Y_{it}^*$$

$$(7) \qquad \text{Liability}_t = \sum_{i=1}^{NF} P_g 0.65 APH_{it}$$

---

[3] A detailed discussion of how augmenting county-level rates with farm-level information (i.e., yields or other individual characteristics) can improve rate accuracy can be found in Rejesus et al. (2006, pp. 410–412).

$$(8) \qquad CPR_t = \frac{\text{Indemnity}_t}{\text{Liability}_t}.$$

Hence, the simulated *CPR* using the *SS* observations in the sample is

$$(9) \qquad CPR = \frac{1}{SS}\sum_{t=1}^{SS} CPR_t.$$

(Milliman and Robertson, Inc. 2000).

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

(10)

$$ER_i = P_g \times 0.65 \times APH_{iSS} \times CPR \times \left(\frac{APH_{iSS}}{Yavc}\right)^{\text{Exp}} + \varepsilon_i,$$

where ε_i is the error term. NLLS was chosen because several ER_i estimates were equal to zero, which makes it impossible to linearize expression using logs. Note that the actual method used by the RMA to calculate the *Exponential* is not publicly available. The only RMA document where exponentials are calculated is Knight's (2000) examination of yield span adjustments, in which an equation similar to equation (10) is estimated through the two-step Heckman procedure. The NLLS method used in this research is consistent with the approach to updating exponents recommended by Coble et al. (2009).

Comparison of True and Estimated Premiums

The premiums estimated through the five procedures (M1, M2, ... , M5) are then compared with the true rates on the basis of 50 runs[4] for each unit of analysis (i.e., NF-SS-CC combination). The statistics used for these comparisons are:

(i) Mean absolute error of the estimated premiums at the farm level (MAD):

---

[4] The decision to make only 50 runs per model was based on time limitations as it took four state-of-the-art computers continuously running for about two months to complete the required grand total of 22,500 runs.

$$(11) \qquad \left(\frac{1}{50 \times NF}\right)\sum_{r=1}^{50}\sum_{j=1}^{NF}\left|\hat{P}_{rj} - \text{Ptrue}_j\right|,$$

where $\hat{P}_{rj}$ is the estimated premium for farm j in run r, and Ptrue$_j$ is the corresponding true premium value. This statistic measures the accuracy of the estimated rates at the farm level.

(ii) The average difference between the estimated and the true premiums (Bias):

$$(12) \qquad \left|\begin{array}{l}\left(\dfrac{1}{50 \times NF}\right)\sum_{r=1}^{50}\sum_{j=1}^{NF}\hat{P}_{rj} \\[2mm] -\left(\dfrac{1}{50 \times NF}\right)\sum_{r=1}^{50}\sum_{j=1}^{NF}\text{Ptrue}_j\end{array}\right|.$$

The premium Bias statistic is useful to ascertain whether, under each particular estimation method, the average premium collections would equal the average indemnities paid.

Results

MAD and Bias Relationships

MAD and Bias measures are computed from the yield simulation and consequent premium estimation results for each of the five distributions in the analysis. To facilitate comparisons, all premiums (true and estimated) are multiplicatively scaled to achieve a true premium average of 10 before calculating the MAD and Bias measures. For each distribution, a total of $3 \times 3 \times 2 \times 5 = 90$ MAD and Bias values are computed on the basis of the 50 runs corresponding to each particular combination of SS, NF, CC, and premium estimation procedures. The relationships between MAD and Bias and the five rate computation procedures for different SS, NF, and CC combinations are then quantified using regression models of the following form:

(13)

$$y = \beta_0 + \beta_{SS}\ln(SS) + \beta_{NF}\ln(NF) + \beta_{CC}CC$$

$$+ \sum_{j=2}^{5}\beta_{M_j}D_{M_j} + \sum_{j=2}^{5}\beta_{M_j,SS}D_{M_j}\ln(SS)$$

$$+ \sum_{j=2}^{5}\beta_{M_j,NF}D_{M_j}\ln(NF) + \sum_{j=2}^{5}\beta_{M_j,CC}D_{M_j}CC + \varepsilon,$$

where (y) is Farm MAD or Bias and the β's are the parameters associated with the explanatory variables, which are the natural logarithm of SS, the natural logarithm of NF, CC, dummy variables for each estimation procedure (D_{Mj}, $j = 2, \ldots 5$), and simple interactions between ln(SS), ln(NF), CC, and the method. A linear-logarithmic model specification was selected because it seemed to provide a better fit than the linear and double-logarithmic models. The dummy variable corresponding to procedure 1 (D_{M1}) is excluded to avoid perfect multicollinearity.

Separate models are estimated for each of the distributions by Ordinary Least Squares (OLS), which results in a total of 10 regressions (two measures of accuracy times five distributions). Standard errors are estimated using the White heteroskedastic-consistent covariance matrix. The use of dummy variables allows for the estimation of premium estimation procedure-specific intercept and slope coefficients. Hence, the full models [equation (13)] are used as the basis for specifying and testing restricted models in which some of the intercepts and slope/interaction parameters are equal across methods and/or set to zero. F tests are conducted to confirm that the set of parameter restrictions imposed in each of the 10 final models are statistically valid. The parameter estimates in the final models and their corresponding covariance matrices are then utilized to estimate and ascertain the statistical significance of the intercepts ($\beta_0 + \beta_{M_j}$, $j = 2, \ldots, 5$) and slope coefficients [$(\beta_{SS} + \beta_{M_j,SS})$, $(\beta_{NF} + \beta_{M_j,NF})$, and $(\beta_{CC} + \beta_{M_j,CC})$, $j = 2, \ldots, 5$] for each premium estimation method.

The parameter estimates and related statistics from the regression models for the two dependent variables of interest (MAD and Bias) under the five assumed yield distributions are available from the authors upon request. In summary, the lowest R^2 value of the 10 models is 0.80, and only two exhibit values lower than 0.90, indicating that a high percentage of the observed variability in the dependent variables is explained by SS, NF, and CC. In addition, nearly 80 percent of the parameters are statistically significant at the 1 percent level, and close to 90 percent at the 10 percent level, suggesting that these variables do in most cases have an impact on MAD and Bias.

Comparison Using Predicted Accuracy and Bias Measures

The previously discussed regression models are used to predict MAD and Bias values across SS, NF, CC, the five premium estimation procedures, and the five assumed yield distributions. The predictions from the final restricted models are preferred to the actual MAD and Bias measures computed directly from the simulation results because they eliminate non-systematic, statistically insignificant variation in the measures across SS, NF, CC, and rate estimation method. Therefore, the observed differences in the predicted measures can be considered statistically valid. Since the five assumed data-generating processes were selected to be representative of the variety of yield distributions that can be encountered in practice, the results are analyzed on the basis of averages across the five distributions.

In addition, to facilitate interpretation, the statistics in Tables 1 and 2 are presented on a percentage basis. For example, the 80.5 MAD value in the first row and column of Table 1 indicates that when using the M1 method, with 25 farm units (NF = 25) in the county, a sample size of 10 observations per farm (SS = 10), and 0 correlation between farm yields in the group (CC = 0), the estimated rates are, on average, 80.5 percent above or below their true values. The 25.3 Bias value displayed in the same location of Table 2 suggests that, under that same scenario, premium estimates exhibit an upward bias of 25.3 percent.

In regard to the MAD, note that even when using the most precise estimation method, percentage deviations can exceed 50 percent in some cases. The reason for this apparently extreme result is that the accuracy of the premium estimate is highly influenced by the precision with which the yield distribution is modeled. Specifically, the rate estimate is very sensitive to differences in the location of the far-most left tail of the estimated yield distribution on which it is based. Since it is difficult to obtain a precise estimate of the far-left tail of the true underlying yield distribution (by either parametric or non-parametric procedures) with a limited number of observations, such high level of premium estimation inaccuracy is not unrealistic. Table 1 also provides insights on how premium estimation accuracy is affected by sample size (SS), the number of farms in the aggre-

Table 1. Percentage MAD of the Premium Estimates at the Farm Level (five-distribution averages)

| | | Correlation Coefficient | | | | | | | | | |
| | | 0 | | | | | 0.5 | | | | |
SS	NF	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
	25	80.5	42.5	63.6	120.8	71.2	76.5	54.6	69.8	122.3	84.5
10	50	78.1	40.5	61.6	120.3	68.7	74.1	52.6	67.9	121.9	82.0
	100	75.6	38.6	59.7	119.8	66.2	71.6	50.8	66.0	121.4	79.5
	25	52.9	35.8	43.7	88.1	50.2	48.9	47.9	49.9	89.7	63.4
25	50	50.5	34.0	41.8	87.7	47.7	46.5	46.0	48.0	89.2	60.9
	100	48.0	32.0	39.9	87.2	45.2	44.0	44.1	46.1	88.8	58.4
	25	32.1	30.9	28.6	63.4	34.3	28.1	43.0	34.8	65.0	47.5
50	50	29.6	29.0	26.7	63.0	31.8	25.6	41.1	33.0	64.6	45.0
	100	27.2	27.1	24.8	62.5	29.3	23.2	39.2	31.0	64.1	42.5
	25	55.2	36.4	45.3	90.8	51.9	51.2	48.5	51.5	92.4	65.1
All	50	52.7	34.5	43.4	90.3	49.4	48.7	46.6	49.6	91.9	62.7
	100	50.3	32.6	41.5	89.8	46.9	46.3	44.7	47.7	91.4	60.1
10		78.1	40.5	61.6	120.3	68.7	74.1	52.7	67.9	121.9	82.0
25	All	50.5	33.9	41.8	87.7	47.7	46.5	46.0	48.0	89.2	60.9
50		29.6	29.0	26.7	63.0	31.8	25.6	41.1	32.9	64.6	45.0
All	All	52.7	34.5	43.4	90.3	49.4	48.7	46.6	49.6	91.9	62.6

Notes:

SS: Sample size.
NF: Number of farms.
M1: Premium estimates obtained from yield distribution models based on the correct distribution, specified with separate means and variances for each farm in the group.
M2: Premium estimates obtained from yield distribution models based on the correct distribution, specified with the same mean and variance for all farms in the group.
M3: Premium estimates obtained from yield distribution models based on the correct distribution, specified with different means but the same variance for all farms in the group.
M4: Premium estimates based on individual farm indemnities using APH yields.
M5: Premium estimates based on RMA-like procedure relying on "group" indemnities and liabilities using APH yields and an exponential correction.

gate rating unit (NF), the correlation coefficient between their yields (CC), and the estimation method (M). An increased NF generally improves accuracy (i.e., decreases the MAD), but only by a small margin (rows All-25, All-50, and All-100). A larger SS substantially improves accuracy in all methods (rows 10-All, 25-All, and 50-All), while a higher correlation coefficient has a mixed and relatively minor effect on accuracy.

When SS=10 (row 10-All), M2 is by far the most accurate method for farm-level premium estimation, but its MAD (40.5 percent when CC=0 and 52.7 percent when CC=0.5) is still substantial. M3 is next best (MAD of 61.6 percent when CC=0 and 67.9 percent when CC=0.5), followed by M1 and M5, while M4 is by far the most inaccurate procedure under this small sample size. At SS=25 (row 25-All), M1 and M3 begin to gain on M2, as an increased sample size improves accuracy the most for those two methods, but, with MADs of 33.9 percent (CC=0) and 46.0 percent (CC=0.5), M2 is still slightly ahead.

Table 2. Percentage Absolute Bias of the Premium Estimates (five-distribution averages)

		Correlation Coefficient									
		0					0.5				
SS	NF	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
	25	25.3	18.2	40.0	11.6	34.0	17.0	20.2	31.8	14.8	32.6
10	50	21.6	18.4	41.8	11.3	32.9	15.6	20.5	33.6	14.5	31.5
	100	20.6	18.7	43.6	11.0	31.9	15.8	20.7	35.4	14.2	30.5
	25	16.2	17.9	21.3	8.2	23.0	8.0	19.9	13.1	11.4	21.6
25	50	12.5	18.1	23.1	7.9	21.9	7.9	20.2	14.9	11.1	20.5
	100	11.5	18.4	24.9	7.6	20.9	8.8	20.4	16.7	10.8	19.5
	25	9.3	17.7	11.7	5.6	14.7	4.7	19.7	7.6	8.8	13.3
50	50	8.5	17.9	9.4	6.1	13.6	4.2	20.0	5.3	8.5	12.2
	100	11.7	18.2	10.8	6.9	12.6	8.2	20.2	4.2	8.2	11.2
	25	16.9	17.9	24.3	8.5	23.9	9.9	20.0	17.5	11.7	22.5
All	50	14.2	18.2	24.8	8.4	22.8	9.2	20.2	17.9	11.4	21.4
	100	14.6	18.4	26.5	8.5	21.8	10.9	20.4	18.8	11.1	20.4
10		22.5	18.4	41.8	11.3	32.9	16.2	20.5	33.6	14.5	31.5
25	All	13.4	18.1	23.1	7.9	21.9	8.2	20.2	14.9	11.1	20.5
50		9.8	17.9	10.6	6.2	13.6	5.7	20.0	5.7	8.5	12.2
All	All	15.2	18.2	25.2	8.4	22.8	10.0	20.2	18.1	11.4	21.4

Notes:

SS: Sample size.

NF: Number of farms.

M1: Premium estimates obtained from yield distribution models based on the correct distribution, specified with separate means and variances for each farm in the group.

M2: Premium estimates obtained from yield distribution models based on the correct distribution, specified with the same mean and variance for all farms in the group.

M3: Premium estimates obtained from yield distribution models based on the correct distribution, specified with different means but the same variance for all farms in the group.

M4: Premium estimates based on individual farm indemnities using APH yields.

M5: Premium estimates based on RMA-like procedure relying on "group" indemnities and liabilities using APH yields and an exponential correction.

Since the biases can be positive or negative depending on the method and the assumed distribution, the figures in the table are computed as the average of the absolute value of the biases across the five distributions.

When SS = 50 (row 50-All), M1 and M3 achieve MADs of under 30 percent and become generally more accurate than M2. Even at this largest SS, the RMA-like historical loss-cost procedures (with the exception of M5 at CC = 0, which performs nearly as well as the parametric methods) still exhibit much higher MADs.

The superior performance of M2 relative to M1 when the sample size is small is explained by the fact that, when using M1, the estimates for the NF individual farm yield means and standard deviations are highly imprecise, often falling far outside the ranges established for the true parameters (140 to 180 bushels per acre and 25 to 35 bushels per acre, respectively). Alternatively, under M2, the average mean and variance estimates (which apply to each and all NF farms) are often close to the true averages (160 and 30) and well within those ranges. Thus, the average estimates from M2 are often closer to the true underlying mean

and variance values than the individual farm-level estimates from M1, and the premiums estimated on the basis of M2 are more accurate. Since M3 estimates have different means but the same variance across all farms in the group, it performs better than M1 but worse than M2 when SS = 10. As the sample size increases, the precision of the M1 estimates for the individual farm yield means and standard deviations improves to where, at SS = 50, they become better than the M2 averages and therefore result in more accurate premium estimates.

In summary, between the two historical loss-cost methods, M5 appears to be preferred for farm-level premium estimation. This is consistent with the recent rating review conducted by Coble et al. (2009), which supports the use of a historical loss-cost procedure akin to M5 where a baseline county rate is augmented by farm-level data. Of the procedures based on parametric yield distribution models (M1, M2, and M3), M2 performs best when SS = 10 and 25, while M1 or M3 are equally preferred for SS = 50. When using the best parametric distribution method for each sample size, substantial gains in farm-level estimation accuracy are observed in comparison to M5, particularly under small sample size conditions, which are more realistic in crop insurance premium rate-setting (Table 1).

In regard to premium estimation bias, since the biases can be positive or negative depending on the method and the assumed distribution, the figures in Table 2 are computed on the basis of the average of the absolute values of the biases across the five distributions. From this table, it is clear that NF has no discernible effect in any of the methods (rows All-25, All-50, and All-100). An increased SS clearly reduces absolute bias in M1, M3, M4, and M5, but appears to have no effect in M2 (rows 10-All, 25-All, and 50-All).

A higher correlation coefficient substantially reduces bias in M1 and M3, has small mixed effects in the case of M2 and M5, and markedly increases it for M4 (row All-All). The latter might be due to the fact that, unlike M1 and M3, the individual farm indemnity method (M4) does not incorporate information about yield correlation across farms. At the smallest SS of 10 (row 10-All), the lowest absolute biases are related to M4 (11.3 percent when CC = 0 and 14.4 percent when CC = 0.5), followed by M1 and M2 (about 19 percent). For SS = 25 (row 25-All), M1 and M4 show

similar biases in the 8 to 15 percent range, while M2 and M3 exhibit overall biases of near 20 percent. When SS = 50 (row 50-All), M1, M3, and M4 perform equally well, achieving an average bias of 8 percent.

In short, depending on the procedure and sample size utilized, the RMA should generally expect low to moderate (10–30 percent) levels of bias in premium estimation (Table 2). It is also noted that while M1, M3, M4, and M5 appear to be consistent estimation methods (i.e., their bias decreases with SS), M2 does not (Table 2). A final but important point in regard to bias is that, in all methods, its magnitude and direction depends on the shape characteristics of the underlying yield distribution. That is, the same procedure can result in a positive bias under one distribution and a negative under another. Since yields have different shape characteristics depending on the region and crop, this could explain the RMA's difficulties in achieving actuarial fairness even at the crop and regional levels.

Impact of Premium Estimation Inaccuracy

As suggested in the introduction, through adverse selection, premium estimation inaccuracy can have a substantial impact on producer participation and the actuarial performance of the crop insurance program. Although this issue is too important and complex to be fully treated in the remainder of this paper, some general preliminary findings are advanced. These findings are under two simplifying and admittedly questionable assumptions: (i) that the farmers are risk-neutral, and (ii) that they know what the actuarially fair premium is. The second assumption is at least in part justified by the fact that the farmer is deeply familiar with his or her entire historical yield experience and the factors that might have positively or negatively affected it in particular years, as well as with the experiences of neighboring farmers with similar production systems and conditions. As a result, he can estimate his true premium much more accurately than can the RMA. In addition, it is assumed that the producers make a very conscientious business decision on whether or not to purchase crop insurance, i.e., that they do not participate if the premium charged by the RMA exceeds their perceived true premium.

If, for example, the distribution of the RMA premium estimates is symmetric and centered at

the true premium, only 50 percent of the farmers will face rates that are lower than what they know to be actuarially fair and thus participate in the program. As a result, the actuarially fair premium corresponding to all participating farmers will be more than or equal to what they are being charged, which means that the program will generate a net loss and have to be subsidized. In addition, the RMA would need to charge a fraction of its estimated premiums if it wants more than 50 percent participation, creating even larger program losses. The premium reductions and extra subsidies required to achieve various participation rates can be computed on the basis of the true premiums and the probability distributions of the premium estimates implied by each of the empirically grounded yield distributions selected for this research under any particular estimation method.

Specifically, for each farm, the true premium (TP) is compared with the subsidized premium (SP), where SP is computed by multiplying the actual premium estimate times the fraction of it to be charged to the producer (FCH), which is one minus the subsidy rate (SR). Given the previously stated assumptions, if TP > SP, the farmer participates in the program. Otherwise, he does not. The program loss-ratio (LR) is the ratio of the indemnities to be paid to the premiums to be collected, i.e., the sum of the true actuarially fair premiums (TP) divided by the sum of the subsidized premiums paid (SP) across all participating producers. Note that, since there are 50 runs for each SS-NF-CC unit of analysis, and NF = 50 in this case, the resulting statistics are based on a total of 2,500 TP vs. SP comparisons per distribution (Table 3).

The relative cost of increasing program participation due to both the higher subsidy rate required and the larger percentage of producers participating is computed as well, by multiplying the percentage of the producers participating in the program (%Part) times one minus the reciprocal of the loss-ratio (%Part[1−(1/LR)]). This relative value is standardized to equal one when the estimated rates are not subsidized (i.e., FCH = 100). Such statistics are computed for all five distributions using the most accurate historical loss-cost procedure for farm-level premium estimation (M5) and assuming the most likely empirical scenario (SS = 10, NF = 50, and CC = 0.5).

First note that, on average across the five distributions, if farmers are charged 100 percent of the estimated premium, less than 50 percent would participate in the program (Table 3). This participation rate, however, ranges from 29 percent to 65 percent depending on the distribution. Even though the rates themselves are not being subsidized, the program loss-ratio is near two due to the fact that all participating producers are paying less than the actuarially fair premium. This, of course, means that the taxpayers (through the federal government) are subsidizing about half of the total program cost. At a 20 percent rate subsidy level (i.e., charging farmers only 80 percent of the estimated premium), overall participation increases to an average of only 58 percent and is still under 50 percent in two of the five cases. Even with a 40 percent subsidy, average participation is just over 70 percent. In this last scenario, the overall loss-ratio increases to 2.3 (i.e., by about 15 percent) but the dollar amount of government subsidies required doubles due to the increased participation rate.

Although these results are based on representative yield distributions for only one crop (corn) and state (Illinois), they are remarkably in line with the participation rates, loss-ratios, and government subsidy figures that have been observed country-wide during the past 10 years.[5] This suggests that the high degree of inaccuracy in the estimation of crop insurance premiums through historical loss-cost procedures identified in this paper could be the key factor behind the need for substantial government subsidies to keep the program solvent.

Conclusions and Recommendations

The main contributions of this paper are to offer and apply a methodology to quantify the levels of inaccuracy under different premium estimation procedures, sample sizes, and other relevant field conditions, and to preliminarily explore the potential improvements in precision that could be expected from applying alternative procedures and longer sample sizes for crop insurance premium estimation. It is concluded that the performance of the loss-cost procedures leaves much to be desired.

---

[5] See http://www.rma.usda.gov/data/sob.html.

Table 3. Predicted Program Participation Rates, Loss-Ratios, and Relative Indemnity-Related Program Costs at Different Premium Subsidy Levels across the Five Selected Yield Distributions under Method 5 (M5), SS = 10, NF = 50, and CC = 0.5

FCH = 100%	Normal	S_{UA}	S_{UB}	S_{BA}	S_{BB}	Average
%Part	33.080	65.040	56.840	28.960	54.320	47.65
LR	1.577	2.809	1.949	1.351	2.222	1.98
Rel. Cost	1.000	1.000	1.000	1.000	1.000	1.00
FCH = 80%	Normal	S_{UA}	S_{UB}	S_{BA}	S_{BB}	Average
%Part	44.880	70.520	68.000	45.160	62.160	58.14
LR	1.653	3.012	2.041	1.449	2.331	2.10
Rel. Cost	1.463	1.124	1.253	1.867	1.191	1.38
FCH = 60%	Normal	S_{UA}	S_{UB}	S_{BA}	S_{BB}	Average
%Part	62.760	77.640	80.040	63.840	71.440	71.14
LR	1.799	3.268	2.273	1.667	2.545	2.31
Rel. Cost	2.306	1.286	1.617	3.400	1.456	2.01

Note: FCH is the fraction of the estimated premium that is being charged to producers (i.e., one minus the subsidy rate), %Part is the percentage of farmers that participate in the program, LR is the loss-ratio (i.e., the ratio of the indemnities to be paid to the total premiums to be collected), and Rel. Cost is the relative cost of increasing program participation due to both the lower FCH required and the larger percentage of producers participating, standardized to be 1.00 when FCH = 100 percent.

While the methods based on estimating yield distributions are found to be substantially more accurate, margins of error in premium estimation remain high, especially at the smaller sample sizes. Still, the numerical estimates provided in this study suggest that using these alternative procedures combined with larger sample sizes could markedly improve the status quo. However, it is recognized that the RMA might be constrained from using such methods given the multiple legislatively imposed objectives and the various types of crops and risks that the program is mandated to cover (Coble et al. 2009), and that implementation of more complex premium estimation procedures could be costly and time-consuming. The sample size result is important considering that current RMA protocol limits the maximum number of observations to be used for rate-setting to 10 (Carriquiry, Babcock, and Hart 2008).

Although the analyses are based on two simplifying and admittedly strong and questionable assumptions, they also suggest that the high degree of inaccuracy in the estimation of crop insurance premiums through historical loss-cost procedures identified in this paper could, by itself, account for most of the substantial subsidies that are being

needed to keep the program solvent. It is important for the RMA, policymakers, farmers, and taxpayers to know just how imprecise the premium estimates are and to be aware of the likely consequences of this inaccuracy. In addition to increasing government costs, vastly incorrect rates can affect the incentives and returns to the private insurance companies that sell federal crop insurance at those rates. Because of these reasons, it is obvious that improving its rate-making procedures should be a top priority for the RMA. As previously mentioned, marked improvement could be accomplished through the combined use of alternative premium estimation methods and larger yield sample sizes.

Therefore, additional research is recommended to ascertain the impact that such potential improvements in premium estimation accuracy on the overall performance of the U.S. crop insurance program could have. Specifically, given reliable farm-level yield data from the main crops and regions, the methodology advanced in this study could be expanded to estimate the effect of various levels of rating improvement on producer participation and indemnity-related program cost to the taxpayers assuming different levels of pro-

ducers' risk aversion and uncertainty in the knowledge of what their actuarially fair premium is. This would allow the RMA to better assess the potential costs and benefits of adopting more precise (but resource-consuming) rating protocols and procedures in relation to the status quo.

References

Barnett, B.J. 2000. "The U.S. Federal Crop Insurance Program." *Canadian Journal of Agricultural Economics* 48(4): 539–551.

Carriquiry, M.A., B.A. Babcock, and C.E. Hart. 2008. "Using a Farmer's Beta for Improved Estimation of Expected Yields." *Journal of Agricultural and Resource Economics* 33(1): 52–68.

Coble, K.H., T.O. Knight, B.K. Goodwin, M.F. Miller, and R.M. Rejesus. 2009. "A Comprehensive Review of the RMA APH and COMBO Rating Methodology: Draft Final Report." Actuarial publication submitted to the USDA Risk Management Agency. Available at http://www.rma.usda.gov/pubs/2009/comprehensivereview.pdf (accessed Fall 2009).

Goodwin, B.K. 1994. "Premium Rate Determination in the Federal Crop Insurance Program: What Do Averages Have to Say About Risk?" *Journal of Agricultural and Resource Economics* 19(2): 382–395.

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, MS.

Harwood, J., R. Heifner, K. Coble, J. Perry, and A. Somwaru. 1999. "Managing Risk in Farming: Concepts, Research, and Analysis." Agricultural Economics Report No. 774, Economic Research Service, U.S. Department of Agriculture, Washington, D.C.

Johnson, N.L. 1949. "System of Frequency Curves Generated by Method of Translation." *Biometrika* 36(1/2): 149–176.

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

Ker, A.P., and B.K. Goodwin. 2000. "Nonparametric Estimation of Crop Insurance Rates Revisited." *American Journal of Agricultural Economics* 82(2): 463–478.

Knight, T.O. 2000. "Examination of Appropriate Yield Span Adjustments by Crop and Region." Report prepared for the Economic Research Service, U.S. Department of Agriculture, Washington, D.C.

Lu, Y., O.A. Ramirez, R.M. Rejesus, T.O. Knight, and B.J. Sherrick. 2008. "Empirically Evaluating the Flexibility of the Johnson Family of Distributions: A Crop Insurance Application." *Agricultural and Resource Economics Review* 37(1): 79–91.

Milliman and Robertson, Inc. 2000. "Actuarial Documentation of Multiple Peril Crop Insurance Ratemaking Procedures." Consulting report prepared for the Risk Management Agency, U.S. Department of Agriculture, Kansas City, MO.

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. 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.

Ramirez, O.A., and T. McDonald. 2006a. "Ranking Crop Yield Models: A Comment." *American Journal of Agricultural Economics* 88(4): 1105–1110.

____. 2006b. "The Expanded and Re-Parameterized Johnson System: A Most Flexible Crop-Yield Distribution Model." Paper presented at the annual meetings of the American Agricultural Economics Association, Long Beach, CA (July 23–26, 2006). Available at http://agecon.lib.umn.edu/ (accessed Fall 2009).

Ramirez, O.A., T.U. McDonald, and C.E. Carpio. 2010. "A Flexible Parametric Family for the Modeling and Simulation of Yield Distributions." *Journal of Agricultural and Applied Economics* 42(2): 1–17.

Ramirez, O.A., S.K. Misra, and J.E. Field. 2003. "Crop Yield Distributions Revisited." *American Journal of Agricultural Economics* 85(1): 108–120.

Rejesus, R.M., K.H. Coble, T.O. Knight, and Y. Jin. 2006. "Developing Experience-Based Premium Discounts in Crop Insurance." *American Journal of Agricultural Economics* 88(2): 409–419.

Sherrick, B.J., F.C. Zanini, G.D. Schnitkey, and S.H. Irwin. 2004. "Crop Insurance Valuation Under Alternative Yield Distributions." *American Journal of Agricultural Economics* 86(2): 406–419.

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(3): 653–659.

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