



*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

*No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.*

# Empirically Evaluating the Flexibility of the Johnson Family of Distributions: A Crop Insurance Application

Yue Lu, Octavio A. Ramirez, Roderick M. Rejesus, Thomas O. Knight, and Bruce J. Sherrick

This article examines the flexibility of the Johnson system of distributions by assessing its performance in terms of modeling crop yields for the purpose of setting actuarially fair crop insurance premiums. Using data from corn farms in Illinois coupled with Monte Carlo simulation procedures, we found that average crop insurance premiums computed on the basis of the Johnson system provide reasonably accurate estimates even when the data are normal or come from a non-normal distribution other than the Johnson system (i.e., a beta). These results suggest that there is potential for using the Johnson system to rate previously uninsured crops that do not have historical insurance performance data upon which to base premium calculations.

**Key Words:** crop insurance, crop yield modeling, Johnson distribution, premium rate setting

A number of studies have been undertaken to address the issue of identifying an appropriate probability distribution for crop yield modeling (see, among others, Gallagher 1987, Nelson and Preckel 1989, Moss and Shonkwiler 1993, Ramirez, Moss, and Boggess 1994, Ramirez 1997, Goodwin and Ker 1998). These efforts are driven in part by the fact that an accurate representation of the probability distribution of crop yields is

critical to the measurement of risks in agricultural production and, consequently, in setting crop insurance premiums. Incorrect representation of the probability distribution of crop yields can lead to inaccurate premiums and cause moral hazard or adverse selection in crop insurance (Coble et al. 1997).

Given the importance of crop yield modeling in agricultural risk analysis, a number of different statistical procedures have been developed to address this issue. These procedures fall under three general categories: (i) parametric approaches, (ii) non-parametric approaches, and (iii) semi-parametric approaches. Each of these general approaches has distinct advantages and disadvantages [see Ramirez and McDonald (2006a) for a brief discussion of this issue]. The focus of this paper is on the parametric approach to crop yield modeling, where it is assumed that the stochastic behavior of the underlying variable of interest (i.e., crop yield) can be adequately represented by a particular parametric probability distribution function (pdf). Note that the main drawback of a parametric approach is the potential inference error from using a pdf that is not flexible enough to accurately model crop yield data. On the other hand, the main advantage of the parametric approach is that it performs relatively well in small

---

Yue Lu and Octavio Ramirez are former Graduate Research Assistant and Professor and Head, respectively, in the Department of Agricultural Economics and Agricultural Business at New Mexico State University in Las Cruces, New Mexico. Roderick Rejesus is Assistant Professor in the Department of Agricultural and Resource Economics at North Carolina State University in Raleigh, North Carolina. Thomas Knight is Professor in the Department of Agricultural and Applied Economics at Texas Tech University in Lubbock, Texas. Bruce Sherrick is Professor in the Department of Agricultural and Consumer Economics at the University of Illinois at Urbana-Champaign in Urbana, Illinois.

This paper was presented as a selected paper at the Crop Insurance and Risk Management Workshop, sponsored jointly by the Northeastern Agricultural and Resource Economics Association, the Risk Management Agency, the Farm Foundation, the Food Policy Institute at Rutgers University, and Cornell University, in Rehoboth Beach, Delaware, on June 12–13, 2007. The workshop received financial support from the Northeast Regional Center for Rural Development. The views expressed in this paper are the authors' and do not necessarily represent the policies or views of the sponsoring agencies.

This research was funded in part by the National Research Initiative (NRI) of the Cooperative State Research, Education and Extension Service (CSREES), U.S. Department of Agriculture, Grant No. 2004-35400-14194, and by the Agricultural Experiment Station of New Mexico State University.

sample applications. As such, the main parametric distributions that have been used in previous studies are the gamma (Gallagher 1987), the beta (Nelson and Preckel 1989), the log-normal (Turvey and Baker 1990, Jung and Ramezani 1999, Stokes 2000), the Weibull (Sherrick et al. 2004), and the inverse hyperbolic sine (Ramirez 1997).

Recently, Ramirez and McDonald (2006a and 2006b) introduced an expanded form of the Johnson family of distributions as another alternative parametric approach for modeling crop yields. They argue that, because the Johnson family can accommodate all mean-variance-skewness-kurtosis (MVS-K) combinations that may be theoretically exhibited by a random variable, it should provide a reasonably accurate approach for modeling any yield distribution that might be encountered in practice. This would address the main disadvantage of parametric models cited in the literature, i.e., the lack of flexibility and the associated specification error risk.

However, the flexibility of the Johnson family of distributions has not been empirically examined with regard to its potential contribution for more accurately setting crop insurance premiums. Although there have been studies that evaluated the effects of different parametric distributional assumptions on crop insurance premiums, most of them only compared the premium rates from a particular parametric distribution (i.e., a beta or a gamma) relative to the rates derived from a normal distribution.

Furthermore, for studies that made premium rate comparisons under different parametric yield distributions (for example, Sherrick et al. 2004), none have used the procedures in Ramirez and McDonald (2006b) for proper yield distribution and premium rate comparisons. In fact, Norwood, Roberts, and Lusk (2004, p. 1111) alluded that empirical application of the procedures in Ramirez and McDonald (2006b) would be a good contribution to the literature since it would allow for more valid empirical crop yield model comparisons and, consequently, more accurate evaluation of crop insurance premium rate effects of different parametric distributions.

Therefore, the main objective of this paper is to empirically examine the flexibility of the Johnson family of distributions by assessing its accuracy in the determination of crop insurance premiums when the true underlying distribution of crop

yields does not belong to the Johnson system but instead follows an alternative distribution such as a beta or a normal. To achieve this objective, a multi-year, farm-level dataset from Illinois corn producers is utilized, in conjunction with simulation-based analytical procedures that build upon the theoretical approach in Ramirez and McDonald (2006a and 2006b).

The results of this study have important implications in modeling crop yields for risk analysis in general, and for crop insurance premium setting in particular. This study also contributes to the literature by providing a reasonable approach to the problem of setting crop insurance premiums for crops that have not been previously insured and hence on which there is no historical insurance program performance upon which to base the premium calculations.

Premium-setting procedures for “traditional crops” with established crop insurance coverage are primarily experience-based, i.e., based on the actual loss history of participating producers. However, this approach is not feasible in setting rates for newly insured crops. Therefore, if accurate premium rates can be derived from a flexible parametric yield distribution like the Johnson system, then this distribution would be helpful in rating “non-traditional crops” that have not previously been insurable. Further, providing a reasonable premium rate setting approach for “non-traditional” crops that have not been covered by crop insurance before would be consistent with the Agricultural and Risk Protection Act (ARPA) of 2000, which made increasing the availability of crop insurance instruments for producers of specialty and underserved agricultural commodities a major policy priority.

### **The Expanded Johnson System: A Brief Description**

Unlike other frequently assumed distributions such as the beta and the gamma, the original Johnson system, which includes the  $S_U$ ,  $S_B$ , and  $S_L$  (or log-normal) distributions, exhibits the key property of being able to accommodate any theoretically feasible skewness-kurtosis (S-K) combination (see Johnson, Kotz, and Balakrishnan 1994). In fact, the  $S_U$  and the  $S_B$  alone are sufficient for this purpose, as the  $S_L$  only spans the curvilinear boundary between the  $S_U$  and  $S_B$ . The lower

bound of the  $S_B$  distribution is given by  $K = S^2 - 2$ , which is also the upper bound for the theoretically impossible S-K region. Figure 1 illustrates the different S-K regions covered by each of the three distributions in the Johnson system, as well as by the beta and the gamma.

Note that, in contrast to the  $S_U$  and  $S_B$ , the gamma distribution only spans a curvilinear segment on the upper right quadrant of the S-K plane. Although, as the  $S_L$ , the gamma distribution can be adapted to cover the mirror image of this segment on the upper left quadrant, the combinations of S-K values allowed by it are still very limited. Also note that the gamma segment is the upper boundary of the S-K area covered by the beta distribution. Although the beta covers a significant area of the S-K plane, the  $S_B$  can accommodate all S-K combinations allowed by the beta, while the beta only covers a subset of the S-K area spanned by the  $S_B$ .

In addition to their limited coverage of the S-K plane, the gamma and the beta exhibit the same handicap of the original Johnson system. That is, because they are two-parameter distributions, any particular S-K combination is always arbitrarily associated with a specific set of mean and variance values. To address this problem, Ramirez and McDonald (2006a and 2006b) developed a re-parameterization procedure that allows the Johnson system to accommodate the same skewness-kurtosis (S-K) combinations spanned by the original system in conjunction with any mean and variance. This approach can also be used to re-parameterize the gamma or the beta distribution so that the statistical performance of these two distributions can be fairly compared with the Johnson system and also increase their flexibility.<sup>1</sup>

## Data and Research Methodology

The data used in this study are from the University of Illinois Endowment Farms. The manag-

ers of the Endowment Farms control over 11,000 acres distributed among farms ranging from 40 acres to 1,200 acres. Twenty-six Endowment Farms are located in 12 counties in Illinois. The farms are rented to more than 40 farm operators predominantly under the common practice of 50-50 share rental arrangements. The manner in which the farms are operated is similar to commercial operations in Illinois and provides high quality yield data under accurate and consistent record-keeping practices. Crop yield data from 1959 to 2003 are available for the 26 Endowment Farms, with sample size varying from 20 observations to 45 observations.

Using the approach in Ramirez and McDonald (2006a), we parametrically fit the Johnson system (i.e., the  $S_U$ ,  $S_B$ , and  $S_L$ ), the beta, and the normal distribution to the yield series of each of the 26 farms. The log-likelihood functions that are maximized to obtain estimates for the parameters of the expanded  $S_U$ ,  $S_B$ ,  $S_L$ , and beta distributions are obtained from Ramirez and McDonald (2006a). The best-fitting model is determined by simple comparison of the maximum log-likelihood function values (MLLFV).

The beta distribution is the alternative non-normal distribution used in this comparative evaluation of the Johnson system because, relative to the other parametric distributions used in the past (e.g., the gamma), it covers a substantially wider area of the skewness-kurtosis (S-K) plane (Figure 1). Theoretically, the beta distribution is also not directly related to the Johnson system. In addition, most of the empirical literature in agricultural economics over the past decade has used the beta distribution to model crop yields [see, for example, Babcock, Hart, and Hayes (2004), among others]. Therefore, it is evident that the beta distribution is the best candidate for conducting a comparative evaluation of the Johnson family.

A Constrained Maximum Likelihood (CML) procedure was used to estimate the parameters of the expanded  $S_U$ ,  $S_B$ ,  $S_L$ , and beta distributions for each of the 26 farm yield data series. The mean, variance, skewness, and kurtosis implied by each of the fitted models were computed based on those parameter estimates and the formulas provided by Ramirez and McDonald (2006a). Also, as in Ramirez and McDonald (2006a), the means and standard deviations are specified as second and first degree polynomials of time:

<sup>1</sup> Hereinafter, the re-parameterized distribution using the Ramirez and McDonald (2006b) approach is also referred to as the "expanded" version of the distribution (i.e., expanded Johnson, expanded beta). The words "re-parameterized" and "expanded" are used interchangeably. Please see Ramirez and McDonald (2006a and 2006b) for a detailed, step-by-step discussion of this re-parameterization procedure. The interested reader may also consult Mood, Graybill, and Boes (1974) for more information on this topic.



$$(2) \quad \sigma_t = (Z_t \sigma) = \sigma_0 + \sigma_1 t; t = 1, \dots, T.$$

Of the 26 farm yield series investigated, the best-fitting distribution was found to be  $S_U$  for six farms, the  $S_B$  for seven farms, and the beta for eight farms. The null hypothesis of normality, conducted through likelihood ratio tests (Ramirez

Based on these results, eight farm yield series were selected to empirically evaluate the flexibility of the Johnson family of distributions: two farms for which  $S_U$  was best-fitting (Farms A and R, hereinafter  $S_U^1$  and  $S_U^2$ ), two farms for which  $S_B$  was best-fitting (Farms B and N, hereinafter  $S_B^1$  and  $S_B^2$ ), two farms for which beta was best-fitting (Farms U and V, hereinafter Beta1 and Beta2), and two farms for which normality could not be rejected (Farms F and Z, hereinafter Norm1

<sup>2</sup> Note that these parameters are also estimated and presented in Ramirez and McDonald (2006a). The interested reader is referred to that article for more details about the issues related specifically to the estimation procedures.

**Table 1. Maximum Log-Likelihood Function Values (MLLFVs) of the 26 Illinois Corn Farms Based on the Fitted  $S_U$ ,  $S_B$ ,  $S_L$ , Beta, and Normal Distributions**

Farm Label	Sample Size	$S_U$ MLLFV	$S_B$ MLLFV	Beta MLLFV	Normal MLLFV	LRTS	Best-Fitting Model
A	44	-183.62	-186.67	-187.24	-191.64	16.03 <sup>3</sup>	$S_U$
B	32	-123.81	-123.81	-126.39	-134.94	22.27 <sup>3</sup>	$S_B$
C	44	-186.38	-182.15	-185.00	-187.61	10.91 <sup>3</sup>	$S_B$
D	43	-189.23	-189.39	-189.54	-192.55	6.63 <sup>2</sup>	$S_U$
E	25	-108.09	-108.00	-107.72	-112.23	9.01 <sup>2</sup>	Beta
F	27	-128.31	-127.08	-127.55	-128.98	3.81 <sup>0</sup>	Normal
G	31	-133.58	-133.57	-133.26	-140.68	14.83 <sup>3</sup>	Beta
H	34	-161.15	-160.20	-160.93	-161.80	3.20 <sup>0</sup>	Normal
I	43	-181.27	-184.84	-184.94	-185.62	8.71 <sup>2</sup>	$S_U$
J	32	-145.96	-145.94	-146.56	-149.20	6.53 <sup>2</sup>	$S_B$
K	27	-120.75	-118.66	-118.98	-126.11	14.90 <sup>3</sup>	$S_B$
L	29	-132.56	-132.49	-132.55	-132.56	0.13 <sup>0</sup>	Normal
M	37	-169.08	-169.08	-168.95	-171.97	6.02 <sup>2</sup>	Beta
N	45	-197.46	-195.15	-196.37	-197.47	4.64 <sup>1</sup>	$S_B$
O	42	-189.54	-188.40	-188.55	-194.36	11.92 <sup>3</sup>	$S_B$
P	42	-195.34	-195.28	-195.31	-197.77	4.97 <sup>1</sup>	$S_B$
Q	40	-174.07	-173.55	-172.74	-178.18	10.88 <sup>3</sup>	Beta
R	33	-145.36	-145.47	-145.67	-159.09	9.46 <sup>3</sup>	$S_U$
S	40	-181.77	-182.35	-182.50	-184.12	4.70 <sup>1</sup>	$S_U$
T	29	-131.07	-131.05	-129.44	-133.79	8.69 <sup>2</sup>	Beta
U	44	-201.83	-201.21	-200.55	-204.01	6.91 <sup>2</sup>	Beta
V	29	-131.07	-126.34	-125.69	-131.64	11.91 <sup>3</sup>	Beta
W	29	-127.78	-131.24	-131.20	-132.56	2.71 <sup>0</sup>	Normal
X	20	-93.45	-93.96	-94.00	-98.42	9.94 <sup>3</sup>	$S_U$
Y	29	-135.14	-135.00	-134.35	-136.90	5.08 <sup>3</sup>	Beta
Z	30	-143.92	-143.26	-143.37	-144.92	3.32 <sup>0</sup>	Normal

Note: MLLFV represents the maximum log-likelihood function values of the estimated distribution. LRTS stands for the likelihood ratio test statistic, which compares the non-normal model (with the highest MLLFV) to the normal model. The superscripts 1, 2, and 3 denote rejection of the null hypothesis of normality at the 10 percent, 5 percent, and 1 percent levels of significance, respectively, using the LRTS; the 0 superscript indicates non-rejection at the 10 percent level. If the null hypothesis of normality is rejected at the 10 percent level, the best-fitting model is the one with the highest MLLFV; otherwise, the best-fitting model is the normal distribution.

and Norm2). The location of these selected farms in the S-K plane is shown in Figure 2.

For the purpose of this evaluation, it is assumed that the best-fitting distribution for each of the selected farms is in fact the true underlying data-generating process (DGP).<sup>3</sup> The next step is to simulate 20 datasets (45 observations per dataset) from each of the eight best-fitting distributions.<sup>4</sup> These datasets were simulated so that the 45 observations in each dataset followed the estimated time trends for the means and the variances of the yield distributions. Once the 160 datasets (8 selected farms  $\times$  20 datasets) are simulated,  $S_U$ ,  $S_B$ , beta, and normal distributions are fitted to each of these datasets, resulting in 640 models (160 datasets  $\times$  4 distributions fitted to each).

The actuarially fair premiums (AFPs) for an Actual Production History (APH) insurance plan (in year 45) are calculated at different coverage levels ( $\lambda$ ) on the basis of each of the 640 yield distribution models. For a traditional APH contract that guarantees a percentage or coverage level ( $\lambda_j = 0.65, 0.70, \dots, 0.80, 0.85$ ) of the “historical” yield ( $y^e$ ) in 5 percent increments, an AFP can be defined as the dollar value of the expected loss  $E(L)$ :

$$(3) \quad AFP = E(L) \times P^g,$$

<sup>3</sup> This assumption suggests that the true underlying DGP for all the farms in the data can only be  $S_U$ ,  $S_B$ , beta, or normal. Two reviewers raised the question: what if the true DGP is not one of these four? The ideal (but practically infeasible) way to address this issue is to fit *all* possible parametric distributions and then compare the performance of the Johnson system relative to all the other parametric distributions. Given the practical limitation of doing this, we simply chose the four distributions above to test the core hypothesis about the flexibility of the Johnson family to calculate actuarially fair premiums (vis-à-vis the beta and normal). Therefore, there is no assurance that the results of this paper will apply when comparing the Johnson versus other parametric distributions (i.e., Weibull or gamma). But we feel that the approach in this paper is a reasonable first step to assess the flexibility of the Johnson family in the context of crop insurance premium rate-setting because the distributions we choose can accommodate all theoretically feasible mean-variance-skewness-kurtosis combinations. Furthermore, the beta and the normal are popular distributions used for crop yield modeling in the agricultural economics literature. Exploring the performance of the Johnson family relative to other parametric distributions (other than beta and normal) would certainly be a worthwhile endeavor in the future, but is outside the scope of the current paper.

<sup>4</sup> Investigating the implications of using an alternative number of datasets and/or observations is beyond the scope of this study. However, this is an interesting issue that merits further exploration in the future.

where  $E(L)$  is the magnitude of the expected loss (in bushels/ac in this case) and  $P^g$  is the price election selected, which in this analysis is assumed to be \$2.20/bu.

To calculate the AFP in equation (3) based on the 640 fitted yield distributions, we first use the estimated mean and variance functions in equations (1) and (2) to estimate the mean and the variance for the last year of the series (i.e., year 45). The estimated mean from equation (1) is then regarded as the mean yield ( $y^e$ ), from which a yield guarantee at  $j$  coverage levels can be computed as follows:  $y_j^g = \lambda_j y^e$ . The actual yields for the last year of the series ( $y_i$ ) are then simulated by drawing skewness and kurtosis values one million times from each of the fitted distributions (i.e.,  $i = 1$  to 1 million). The loss magnitude (in bu/ac) for a given actual yield occurrence ( $y_i$ ) can then be defined as

$$(4) \quad L_{ji} = \max[(y_j^g - y_i), 0].$$

The expected loss for any coverage level  $\lambda_j$  is then computed as

$$(5) \quad E(L_j) = \sum_{i=1}^m (L_{ji}) / m,$$

where  $m$  is the number of simulated actual yield observations (one million in this case). From the calculated value in equation (5), we can then compute the AFP at  $j$  different coverage levels for each fitted yield distribution using equation (3). The calculated AFPs are then used to evaluate the economic flexibility of the Johnson family—i.e., how far off the calculated AFP is when the yield distribution is assumed to be  $S_B$  but the actual data-generating process is, say, beta.

## Results and Discussion

### Flexibility Evaluation

The first step of the evaluation is to ascertain the frequency with which the true underlying distribution can be identified using the likelihood ratio test (LRT) and, if normality is rejected ( $\alpha = 0.10$ ), using the maximum log-likelihood function value (MLLFV) criteria to select the best-fitting distri-

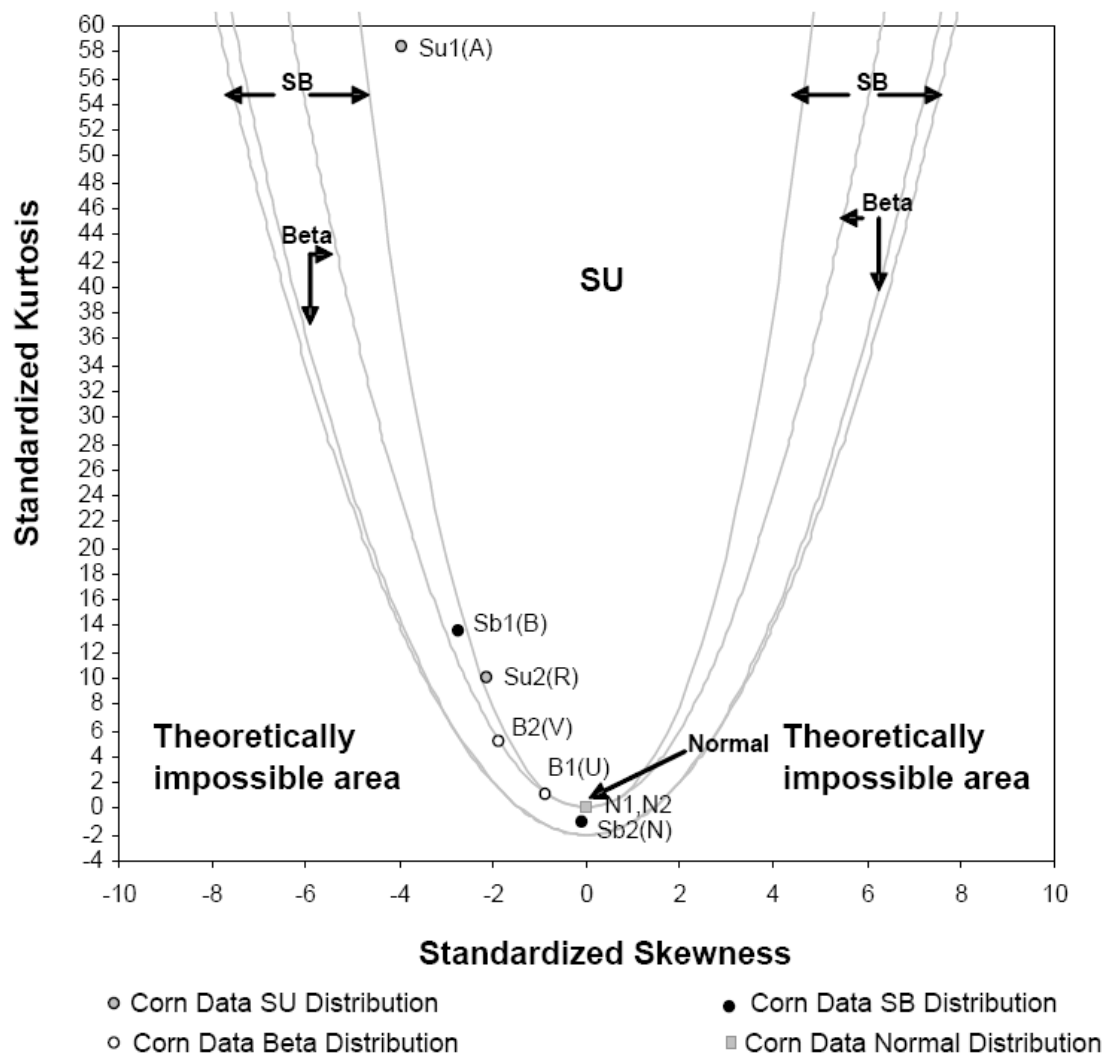


Figure 2. Location of the Selected Farms in the Standardized S-K Plane

bution (Table 2). For example, of the 40 datasets (2 farms  $\times$  20 datasets) for which the  $S_U$  is the true underlying distribution, 30 are correctly identified as  $S_U$ —i.e., normality is rejected—and the  $S_U$  model shows the highest MLLFV in these 30 cases. Of the other 10 cases, three are identified as  $S_B$ , four as beta, and normality cannot be rejected in three cases. Of the 40 simulated datasets that are normal, 34 are correctly identified as normal based on the LRT, i.e., normality is not rejected in 34 cases. These results indicate that for a sample size of 45, when the true distribution is normal or  $S_U$ , the LRT and MLLFV criteria are

able to identify the correct distribution underlying the data (in most cases). This means that in practice, when the true DGP is unknown, but can be reasonably assumed to be one of the four DGPs we consider in this study— $S_U$ ,  $S_B$ , beta, and normal—if it is found that an  $S_U$  distribution best fits the data, one can be reasonably confident that the true DGP is  $S_U$ .

In contrast, when the true underlying distribution is  $S_B$ , only 20 of the datasets are identified as  $S_B$  by the LRT and MLLFV criteria, but, interestingly, the beta exhibits the highest MLLFV in 16 of the other 20 datasets. The remaining 4 cases



**Table 2. Frequency by Which the Underlying Data-Generating Process (DGP) is Identified Using MLLFV Values and LRTs ( $\alpha = 0.10$ )**

Underlying DGP	Number of Times (out of 20) the Distribution Below is Best-Fitting			
	S <sub>U</sub>	S <sub>B</sub>	Beta	Normal
S <sub>U</sub> <sup>1</sup>	16	0	1	3
S <sub>U</sub> <sup>2</sup>	14	3	3	0
S <sub>B</sub> <sup>1</sup>	4	11	5	0
S <sub>B</sub> <sup>2</sup>	0	9	11	0
Beta1	3	6	5	6
Beta2	0	12	8	0
Norm1	0	2	1	17
Norm2	0	1	2	17

are classified as S<sub>U</sub>. Alternatively, when the true underlying distribution is beta, only 13 of the datasets are identified as beta, 18 are identified as S<sub>B</sub>, three as S<sub>U</sub>, and normality is not rejected in six cases. In short, when the true underlying distribution is S<sub>B</sub> or beta, the LRT/MLLFV criteria favors one of these two distributions in most cases.

This preliminary analysis empirically supports the following theoretical hypotheses which are originally based on the regions of the S-K space that these distributions can span:

- The S<sub>B</sub> and the beta are fairly “interchangeable” as probability distribution models; i.e., according to the MLLFV criteria, the S<sub>B</sub> can represent beta DGPs as well as the beta itself, while, in many cases, the beta can be a good model for the S<sub>B</sub> DGPs.
- In most cases, the S<sub>B</sub> and the beta are not “interchangeable” with the S<sub>U</sub> as probability distribution models, and vice-versa.

Additional evidence regarding the flexibility of the S<sub>B</sub> and the beta to substitute for each other as probability distribution function models can be obtained by comparing the MLLFVs, skewness, and kurtosis estimates from the fitted distributions (Table 3).<sup>5</sup> First, for Farms Beta1 and Beta2, no-

tice that the average MLLFV value is actually slightly higher for the S<sub>B</sub> models even though the DGPs in these cases are beta. Although this is obviously a random occurrence (i.e., under a large sample size or on average over a large enough number of samples that the model based on the true distribution will always yield the highest MLLFV), it is an indication that the S<sub>B</sub> model can represent beta DGPs quite well. That is, when the DGP is beta, both models exhibit a similar likelihood of having generated the data.

Alternatively, for Farms S<sub>B</sub><sup>1</sup> and S<sub>B</sub><sup>2</sup> where the S<sub>B</sub> is the underlying DGP, the beta models exhibit noticeably lower MLLFVs than the S<sub>B</sub> models. This suggests that the beta model might not be able to approximate an S<sub>B</sub> DGP as well as the S<sub>B</sub> can represent the beta. In addition, note that the S<sub>U</sub> models, on average, exhibit moderately to substantially smaller MLLFVs than the S<sub>B</sub> and the beta models when the DGP is S<sub>B</sub> or beta, and the S<sub>B</sub> and beta models show markedly lower MLLFVs than the S<sub>U</sub> when the DGP is S<sub>U</sub>.

In general, we also find that the average skewness and kurtosis values implied by the fitted S<sub>B</sub> distributions are closer to the true skewness and kurtosis of the underlying beta DGPs, compared to how close the average skewness and kurtosis from the beta models are to the correct skewness and kurtosis of the underlying S<sub>B</sub> DGPs. Similar to what was observed in the MLLFV comparisons, the S<sub>B</sub> and beta models are better able to estimate the skewness-kurtosis combinations of each other's DGPs than the S<sub>U</sub> can, and the S<sub>U</sub> does a relatively poor job of estimating the skewness and kurtosis of both the S<sub>B</sub> and the beta DGPs.

When taken together, the results from the analyses in this sub-section suggest that the Johnson family (in particular, the S<sub>U</sub> and the S<sub>B</sub> distributions) might be an acceptable approach to modeling crop yields when the true underlying DGP is unknown (but can be reasonably assumed to be one of the four DGPs we consider in this study—S<sub>U</sub>, S<sub>B</sub>, beta, and normal). At the very least, it can be concluded that the Johnson system is more

<sup>5</sup> In the interest of space, only the average MLLFV, skewness, and kurtosis values are reported. Those averages are calculated over the 20 datasets for each of the farms where the true DGPs are S<sub>U</sub>, S<sub>B</sub>, and beta.

Note that the normal is not included because the skewness and kurtosis of this distribution will always be zero and three, respectively. The individual MLLFV, skewness, and kurtosis values for each of the fitted distributions are available from the authors upon request.

**Table 3. Average MLLFV, Skewness, and Kurtosis Values for the Fitted Distributions Calculated Over the 20 Simulated Data Sets for Each Selected Farm**

Selected Farm/Fitted Distribution	Average MLLFV	St. Dev. of MLLFV	Average Skewness	St. Dev. of Skewness	Average Kurtosis	St. Dev. of Kurtosis
A. Farm $S_U^1$						
True	--	--	-3.94	--	58.33	--
$S_U$	-180.49	8.51	-2.28	1.90	35.86	43.81
$S_B$	-185.42	8.13	-0.91	0.61	1.90	2.32
Beta	-185.97	8.30	-0.73	0.44	0.89	3.43
B. Farm $S_U^2$						
True	--	--	-2.10	--	10.01	--
$S_U$	-196.47	7.54	-1.86	0.71	8.22	6.24
$S_B$	-197.89	6.60	-1.26	0.58	3.24	3.81
Beta	-198.25	7.25	-1.14	0.34	2.00	1.14
C. Farm $S_B^1$						
True	--	--	-2.73	--	13.56	--
$S_U$	-173.33	7.28	-4.76	1.99	77.61	80.82
$S_B$	-172.56	7.36	-2.10	1.34	11.38	20.34
Beta	-173.09	7.81	-1.63	0.25	3.75	1.17
D. Farm $S_B^2$						
True	--	--	-0.10	--	-1.13	--
$S_U$	-198.79	4.34	-0.31	0.47	0.58	1.04
$S_B$	-191.64	2.51	-0.08	0.19	-1.27	0.18
Beta	-192.93	4.86	-0.17	0.31	-0.81	0.60
E. Farm Beta1						
True	--	--	-0.87	--	0.98	--
$S_U$	-202.54	5.09	-1.04	0.53	2.73	2.81
$S_B$	-202.20	5.13	-0.72	0.37	0.49	1.29
Beta	-202.21	5.09	-0.77	0.36	0.67	1.08
F. Farm Beta2						
True	--	--	-1.86	--	5.10	--
$S_U$	-195.44	5.64	-5.27	2.44	104.82	97.24
$S_B$	-193.68	5.46	-1.52	0.62	2.92	4.05
Beta	-194.20	5.39	-1.76	0.13	4.51	0.65

Note: The “true” value above is based on the estimated values from the assumed underlying DGP of the farm (i.e., Farm Beta1’s DGP is assumed to be beta with the “true” skewness and kurtosis values reported above).

likely to provide a reasonably accurate representation of an unknown underlying DGP than the beta distribution.

In practice, one can first estimate the expanded  $S_U$  and  $S_B$  models for the particular yield data series of interest. If normality is rejected by the LRT and the  $S_U$  has the highest MLLFV, then based on the previous results one can be fairly confident that the underlying DGP is  $S_U$  rather than  $S_B$  or beta, and that the estimated skewness and kurtosis values will be fairly close to the true ones. In addition, if it is found that the yield series is non-normal and the  $S_B$  has the highest MLLFV, it can be expected that the estimated skewness and kurtosis values are fairly close to the true ones even if the DGP is beta rather than an  $S_B$ . More importantly, it is likely that if a beta model were to be estimated in such an instance, its MLLFV would not be noticeably higher than the  $S_B$ —i.e., a significant improvement in the goodness-of-fit should not be expected.

Overall, the results of this sub-section can be summarized as follows:

- The  $S_B$  is a reasonably close “surrogate” model for the beta, while the beta appears to be less capable of properly substituting for the  $S_B$  (at least in some cases).
- The  $S_B$  and the beta are generally not very good substitutes for the  $S_U$ , and the  $S_U$  is a poor surrogate for either the beta or the  $S_B$ .

Although MLLFV and skewness-kurtosis estimate comparisons clearly support the previously stated conclusions, a final assessment of whether one distribution can be considered a close enough substitute for another can be made only on the basis of whether using the substitute instead of the exact underlying distribution can substantially affect the results of an economic analysis. Such an assessment is conducted in the following sub-section.

#### *Flexibility Assessment Using AFPs*

Further empirical investigation of the flexibility of the Johnson family is undertaken by comparing the AFPs calculated on the basis of the estimated distributions ( $S_U$ ,  $S_B$ , beta, and normal) versus the correct AFP calculated using the true underlying distribution for each of the eight previously discussed cases. At the individual farm level ( $n = 45$

observations), it is observed that using any estimated distribution, even the correct one, will generally lead to imprecise AFPs.<sup>6</sup> It is important to emphasize that this imprecision also holds when the correct estimated distribution is used to compute the premium. In other words, such lack of accuracy appears to be due to the use of single, relatively small samples rather than to the choice of an erroneous distributional model.

The average of the 20 premiums calculated on the basis of the estimated Johnson system (i.e., either the  $S_U$  or the  $S_B$  distribution depending on which exhibits the highest MLLFV)<sup>7</sup> and the true distribution, for eight different coverage levels (0.5, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, and 0.85), are presented in Table 4. First note that the premiums obtained from the estimated Johnson system are quite close to the true premiums regardless of the distribution underlying the DGP. Specifically, in the two cases when the DGP is beta (Beta1 and Beta2), the average of the Johnson system premiums are \$5.12 and \$1.33, respectively, versus the average of the correct premiums for both beta farms of \$5.74 and \$1.34. Notice that the average premiums implied by the estimated beta models for the two Beta farms are \$4.72 and \$1.05, respectively, which is further from the true premiums (Table 4).

For the Johnson system, the average of the absolute differences (AAD) between the estimated and the true premiums across the eight coverage levels is \$0.62 in the case of the first beta DGP (Beta1) and \$0.06 in the case of the second (Beta2). Interestingly, in contrast, the AAD between the premiums computed from the estimated beta models (Table 4) and the true ones are \$1.02 and \$0.29, respectively. That is, on average, the Johnson system does better than the correct model in estimating the true premiums in these two par-

<sup>6</sup> In the interest of space, AFPs at the individual level (640 AFPs) are not reported here but are available from the authors upon request.

<sup>7</sup> This means that the “Johnson system” approach discussed here (and in Table 4) is where we calculate premiums for the individual farms (i.e., one of the 20 for each underlying DGP) using either an  $S_U$  or  $S_B$ , depending on which one had the highest MLLFV. For example, if in the 20 farms/datasets under Beta1 there are eight farms where  $S_U$  had the highest MLLFV and twelve farms where  $S_B$  had the highest MLLFV, then the average premium reported in Table 4 is the average premium based on individual premiums calculated based on eight  $S_U$ -fitted series and twelve  $S_B$ -fitted series. In contrast, an average premium based on, say,  $S_B$  alone, is calculated by using all twenty  $S_B$ -fitted series (regardless of whether or not eight farms showed that an  $S_U$  should be a better fit).

**Table 4. Average AFPs Calculated Over 20 Simulated Data Sets Using the Combination of Best-Fitting Johnson Distributions (i.e.,  $S_U$  or  $S_B$ ) for All the Selected Farms**

Farm/Fitted Distribution	Coverage Level								Avg. Premium	Avg. Abs. \$ Diff.
	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85		
Farm $S_U^1$										
True	1.80	2.14	2.57	3.11	3.82	4.74	5.99	7.72	3.99	--
Johnson	1.87	2.21	2.64	3.21	3.97	5.01	6.45	8.50	4.23	0.24
Beta	0.22	0.33	0.51	0.80	1.26	2.00	3.22	5.18	1.69	2.30
Farm $S_U^2$										
True	1.97	2.44	3.04	3.81	4.82	6.14	7.86	10.14	5.03	--
Johnson	1.58	1.98	2.50	3.18	4.10	5.35	7.05	9.35	4.39	0.64
Beta	0.42	0.64	0.98	1.46	2.18	3.23	4.76	6.95	2.58	2.45
Farm $S_B^1$										
True	0.60	0.84	1.16	1.59	2.18	2.98	4.09	5.65	2.39	--
Johnson	1.10	1.33	1.62	2.00	2.51	3.25	4.35	6.06	2.78	0.39
Beta	0.13	0.20	0.32	0.50	0.80	1.29	2.09	3.44	1.10	1.29
Farm $S_B^2$										
True	0.00	0.00	0.00	0.00	0.06	0.84	2.86	6.28	1.25	--
Johnson	0.00	0.00	0.00	0.00	0.02	0.71	2.94	6.77	1.31	0.09
Beta	0.04	0.07	0.13	0.24	0.44	1.02	2.58	5.53	1.26	0.26
Farm Beta1										
True	1.51	2.12	2.94	4.03	5.47	7.33	9.73	12.77	5.74	--
Johnson	1.44	1.94	2.62	3.53	4.78	6.45	8.66	11.54	5.12	0.62
Beta	1.20	1.67	2.31	3.18	4.36	5.96	8.11	10.95	4.72	1.02
Farm Beta2										
True	0.10	0.17	0.30	0.52	0.90	1.55	2.66	4.55	1.34	--
Johnson	0.16	0.23	0.34	0.53	0.85	1.43	2.53	4.57	1.33	0.06
Beta	0.07	0.12	0.21	0.37	0.66	1.17	2.07	3.69	1.05	0.29
Farm Norm1										
True	0.15	0.32	0.63	1.17	2.09	3.55	5.77	8.97	2.83	--
Johnson	0.09	0.19	0.42	0.91	1.85	3.45	5.91	9.45	2.78	0.05
Beta	0.04	0.10	0.25	0.59	1.31	2.64	4.83	8.10	2.23	0.60
Farm Norm2										
True	0.00	0.01	0.04	0.12	0.33	0.80	1.77	3.58	0.83	--
Johnson	0.01	0.02	0.03	0.09	0.24	0.68	1.71	3.71	0.81	0.02
Beta	0.03	0.05	0.10	0.20	0.46	1.06	2.31	4.55	1.10	0.27

Notes: The average AFP for a farm is calculated by first taking the AFP of each dataset and then taking the average of the 20 calculated AFPs. For the Johnson approach, the AFP for each dataset is calculated based on the best-fitting distribution (either the  $S_U$  or the  $S_B$ , whichever has the highest MLLFV). The average premium is then calculated by taking the average of the AFPs across coverage levels.

“Avg. Abs. \$ Diff.” is the average of the absolute dollar differences over all coverage levels. This is calculated by taking the absolute value of the difference between the true AFP and the estimated AFP for each coverage level and then taking the average over the eight coverage levels.

ticular cases. Although this is not generally to be expected, it indicates that at a sample size of 45 there is no evidence that if the underlying DGP is a beta, using an estimated beta model to compute the premiums would provide more accurate results than using the Johnson system.

The opposite, however, is not necessarily true. That is, when the DGP is the Johnson system, using estimated beta models to compute the premiums can result in considerable imprecision (i.e., differences with the true premiums), even when looking at averages over 20 samples. As expected, the differences are particularly noteworthy when the underlying DGP is  $S_U$ . Specifically, the AAD between the beta-estimated and the true premiums is \$2.30 ( $S_U^1$ ) and \$2.45 ( $S_U^2$ ), versus \$0.24 ( $S_U^1$ ) and \$0.64 ( $S_U^2$ ) when the premiums are computed on the basis of the estimated Johnson system (mostly the estimated  $S_U$  model).<sup>8</sup>

Finally, it is important to note that these premium comparison analyses support the hypothesis that the  $S_B$  is generally better in approximating a beta-generated DGP than the beta is in approximating an  $S_B$  distribution. Specifically, under DGPs  $S_B^1$  and  $S_B^2$ , the AADs between the beta-estimated and the true premiums are \$1.29 and \$0.26, respectively (Table 4). Although not reported in Table 4, we find that when the true DGPs Beta1 and Beta2 are considered, the AADs between an  $S_B$ -estimated premium (i.e., from  $S_B$  alone) and the true premiums are \$1.20 and \$0.06.<sup>9</sup>

<sup>8</sup> Although not presented here (due to space considerations), we also considered the case when the true DGP is  $S_U$  but the AFP is calculated based solely on an  $S_B$ . That is, we don't use the "Johnson system" approach where we calculate the premiums based on either the  $S_U$  or  $S_B$  depending on which had the highest MLLFV (i.e., in the case not presented in the text, even if the  $S_U$  is the best-fitting model in some or all of the 20 datasets for each DGP, we still used the  $S_B$  distribution to calculate the premiums in this situation). In this case, the AAD between the  $S_B$ -generated premium and the true premium (from an  $S_U$ ) is large as well. This supports the notion that the  $S_B$  per se may not be a good surrogate for  $S_U$  when calculating premiums over 20 samples. In addition, although not reported here, we also considered the case of calculating the AFP based solely on a normal distribution when the true DGP is non-normal (i.e., beta or  $S_U$  or  $S_B$ ), and found that AFPs based on the normal distribution alone are also very imprecise. Results from this analysis are available from the authors upon request.

<sup>9</sup> As in footnote 8, we did not explicitly present this in Table 4, but we also considered the case when the true DGP is beta but the AFP is calculated based solely on an  $S_B$ . The calculated AAD for this case is the one reported in the text. Again, results from this analysis are available from the authors upon request.

In short, the previously discussed empirical results support the theory-based hypothesis that the Johnson system is a fairly flexible parametric distribution model and can be used to fairly accurately estimate AFPs for crop insurance when the true underlying distribution of crop yields is unknown. Under such conditions, the results certainly demonstrate that the Johnson system should be preferred to the most commonly used parametric alternative, i.e., the beta distribution. Some measure of error should of course still be expected when using the Johnson system to estimate AFPs, but the magnitude of those errors seems to be reasonable. Specifically, when estimating an average AFP across 20 farm yield series (with 45 observations each), the Johnson system is shown to provide reasonably accurate estimates even if the true underlying distribution is not part of the Johnson family (say, a beta or a normal).

## Conclusions and Implications

This study examines the flexibility of the Johnson system (i.e., the  $S_U$  and  $S_B$  distributions) by evaluating the accuracy of actuarially fair crop insurance premiums estimated on the basis of this system, when the true underlying yield distribution is not within the Johnson system. It is found that average AFPs calculated from the Johnson system are reasonably accurate estimates for the correct rates even when the true distribution underlying the data is a beta or a normal. The results imply that if an average premium rate is needed for a certain geographic area, the Johnson system can be used for reliably estimating this premium, provided that there are a sufficient number of farms to average across.

This may have strong implications in rating previously uninsured crops. Specifically, the average AFP for a particular base area (i.e., a county or a crop-reporting district) can be used as a basis for an Actual Production History (APH) crop insurance program (much like how the reference rate is now used in the current APH rating system). An individual risk classification system together with the average AFP calculated using the Johnson system can be used to generate individual rates.<sup>10</sup>

<sup>10</sup> In the current APH, a yield ratio (the ratio of individual yield and county yield) and an exponential curve is used as the individual risk

One caveat about the results reported in this article is that the crop insurance rating performance of the Johnson system is only evaluated relative to hypothetical scenarios where the underlying yield distributions are beta or normal. Although the results are indicative of the potential of the Johnson system to approximate other parametric distributions for the purpose of crop insurance premium setting, further research is needed using other underlying distributions such as the Weibull, gamma, or more complex normal mixtures in order to fully confirm this potential.

## References:

- Babcock, B.A., C.E. Hart, and D.J. Hayes. 2004. "Actuarial Fairness of Crop Insurance Rates with Constant Rate Relationships." *American Journal of Agricultural Economics* 86(3): 563–575.
- Coble, K.H., T.O. Knight, R.D. Pope, and J.R. Williams. 1997. "An Expected Indemnity Approach to the Measurement of Moral Hazard in Crop Insurance." *American Journal of Agricultural Economics* 79(1): 216–226.
- Gallagher, P. 1987. "U.S. Soybean Yields: Estimation and Forecasting with Nonsymmetric Disturbances." *American Journal of Agricultural Economics* 69(4): 796–803.
- Goodwin, B.K., and A.P. Ker. 1998. "Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts." *American Journal of Agricultural Economics* 80(1): 139–153.
- 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.
- Johnson, J.L., S. Kotz, and N. Balakrishnan. 1994. *Continuous Univariate Distributions*. New York: Wiley and Sons.
- Jung, A.R., and C.A. Ramezani. 1999. "Valuing Risk Management Tools as Complex Derivatives: An Application of Revenue Insurance." *Journal of Financial Engineering* 8(March): 99–120.
- Ker, A.P., and K. Coble. 2003. "Modeling Conditional Yield Densities." *American Journal of Agricultural Economics* 85(2): 291–304.
- Mood, A.M., F.A. Graybill, and D.C. Boes. 1974. *Introduction to the Theory of Statistics* (3rd edition). New York: McGraw-Hill Book.
- Moss, C.B., and J.S. Shonkwiler. 1993. "Estimating Yield Distributions with a Stochastic Trend and Non-Normal Errors." *American Journal of Agricultural Economics* 75(4): 1056–1062.
- Nelson, C.H., and P.V. Preckel. 1989. "The Conditional Beta Distribution as a Stochastic Production Function." *American Journal of Agricultural Economics* 71(2): 370–377.
- Norwood, B., M.C. Roberts, and J.L. Lusk. 2004. "Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions." *American Journal of Agricultural Economics* 86(4): 1032–1043.
- Ramirez, O.A. 1997. "Estimation and Use of a Multivariate Parametric Model for Simulating Heteroscedastic, 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.U. McDonald. 2006a. "The Expanded and Re-parameterized Johnson System: A Most Flexible Crop-Yield Distribution Model." Selected paper at the AAEA Annual Meetings, Long Beach, CA (July 23–26).
- . 2006b. "Ranking Crop Yield Models: A Comment." *American Journal of Agricultural Economics* 88(4): 1105–1110.
- Ramirez, O.A., C.B. Moss, and W.G. Boggess. 1994. "Estimation and Use of the Inverse Hyperbolic Sine Transformation to Model Non-Normal Correlated Random Variables." *Journal of Applied Statistics* 21(4): 289–305.
- 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.
- Stokes, J.R. 2000. "A Derivative Security Approach to Setting Crop Revenue Insurance Premiums." *Journal of Agricultural and Resource Economics* 25(1): 159–176.
- Taylor, C.R. 1990. "Two Practical Procedures for Estimating Multivariate Non-Normal Probability Density Functions." *American Journal of Agricultural Economics* 72(1): 210–217.
- Turvey, C.G., and T.G. Baker. 1990. "A Farm-Level Financial Analysis of Farmers' Use of Futures and Options under Alternative Farm Programs." *American Journal of Agricultural Economics* 72(4): 946–957.

---

classification system to generate individual premiums based on an average AFP or reference rate. A similar system may be used if one is to implement an APH-type crop insurance program for a previously uninsured crop.