



*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.*

# A Flexible Parametric Family for the Modeling and Simulation of Yield Distributions

Octavio A. Ramirez, Tanya U. McDonald, and Carlos E. Carpio

The distributions currently used to model and simulate crop yields are unable to accommodate a substantial subset of the theoretically feasible mean-variance-skewness-kurtosis (MVS<sub>K</sub>) hyperspace. Because these first four central moments are key determinants of shape, the available distributions might not be capable of adequately modeling all yield distributions that could be encountered in practice. This study introduces a system of distributions that can span the entire MVS<sub>K</sub> space and assesses its potential to serve as a more comprehensive parametric crop yield model, improving the breadth of distributional choices available to researchers and the likelihood of formulating proper parametric models.

*Key Words:* risk analysis, parametric methods, yield distributions, yield modeling and simulation, yield nonnormality

**JEL Classifications:** C15, C16, C46, C63

Agricultural economists have long recognized that the choice of an appropriate probability distribution to represent crop yields is critical for an accurate measurement of the risks associated with crop production. Anderson (1974) first emphasized the importance of accounting for nonnormality in crop yield distributions for the purpose of economic risk analysis. Since then, numerous authors have focused on this issue, including Gallagher (1987), Nelson and Preckel (1989), Moss and

Shonkwiler (1993), Ramirez, Moss, and Boggess (1994), Coble et al. (1996), and Ramirez (1997). These authors have provided strong statistical evidence of nonnormality and heteroskedasticity in crop yield distributions—specifically, the existence of kurtosis and negative skewness in a variety of cases. The possibility of positive skewness has been documented as well (Ramírez, Misra, and Field, 2003).

The three general statistical procedures that have been used for the modeling and simulating of crop yield distributions are the parametric, the nonparametric, and semiparametric. All have distinct advantages and disadvantages. The parametric procedures assume that the data-generating process can be adequately represented by a particular parametric probability distribution function. For this reason, the main disadvantage of this method is the potential error resulting from assuming a probability distribution that is not flexible enough to properly represent the data. The main advantage of this method is that, if the assumed distribution can adequately

---

Octavio A. Ramirez is professor and head, Department of Agricultural and Applied Economics, University of Georgia, Athens, GA. Tanya U. McDonald is research specialist, Department of Agricultural Economics and Agricultural Business, New Mexico State University, Las Cruces, NM. Carlos E. Carpio is assistant professor, Department of Applied Economics and Statistics, Clemson University, Clemson, SC.

This study was supported by the National Research Initiative of the Cooperative State Research, Education and Extension Service, USDA, Grant #2004-35400-14194, and by the Agricultural Experiment Stations of the University of Georgia and New Mexico State University.

represent the data-generating process, it performs relatively well even in small sample applications. This is important because crop yield datasets do not often span long periods. Distributions that have been used as a basis for parametric procedures include the Normal, the Log-normal, the Logistic, the Weibull, the Beta, the Gamma, and the Inverse Hyperbolic Sine.

The nonparametric approaches also have strengths and weaknesses. Because these methods are free of a functional form assumption, they are generally more flexible. However, they can be inefficient relative to parametric procedures under certain conditions. Specifically, according to Ker and Coble (2003), "it is possible, perhaps likely, for very small samples such as those corresponding to farm-level yield data, that an incorrect parametric form—say Normal—is more efficient than the standard nonparametric kernel estimator." Other authors cite theoretical complexity and intensive computational requirements as another disadvantage of nonparametric procedures (Yatchew, 1998). Semiparametric methods show significant potential because they encapsulate the advantages of the parametric and nonparametric approaches while mitigating their disadvantages (Ker and Coble, 2003; Norwood, Roberts, and Lusk, 2004). Because the semiparametric approach is based on nonparametrically "correcting" a particular parametric estimate, the availability of more flexible distributions such as the ones advanced in this study should improve the potential efficiency of semiparametric procedures as well.

Extensive efforts have been devoted to the issue of the most appropriate probability distribution to be used as a basis for parametric or semiparametric methods. Gallagher (1987) used the well-known Gamma density as a parametric model for soybean yields. Nelson and Preckel (1989) proposed a conditional Beta distribution to model corn yields. Taylor (1990) estimated multivariate nonnormal densities through a conditional distribution approach based on the Hyperbolic Tangent transformation. Ramirez (1997) introduced a modified Inverse Hyperbolic Sine transformation (also known in the statistics literature as the  $S_U$  Distribution) as a possible multivariate nonnormal and heteroskedastic crop yield distribution model. Ker and

Coble (2003) proposed a semiparametric model based on the Normal and Beta densities. Empirical comparisons of leading parametric models have been recently attempted (Norwood, Roberts, and Lusk, 2004). Despite such advances, the potential of different probability density functions (pdfs) to serve as suitable yield and price distribution models has not been assessed in the context of a rigorous theoretical framework, and this critical research methods issue remains unsettled.

According to basic statistical theory (Mood, Graybill, and Boes, 1974), the first four central moments of a pdf are the main descriptors of its shape. Although there are other means for characterizing and comparing distributions, this suggests that the flexibility of a pdf to accommodate a variety of empirical shapes (i.e., data-generating processes) should be closely related to the Mean-Variance-Skewness-Kurtosis (MVSK) combinations that are allowed by it.

Unfortunately, in this regard, all of the pdfs that have been used as a basis for parametric methods suffer from two significant range restrictions: (1) they can only accommodate limited subsets of all of the theoretically feasible Skewness-Kurtosis (SK) combinations and, therefore, might only be capable of adequately modeling underlying data distributions where third and fourth central moments are within that subset; and (2) their variance, skewness, and kurtosis are controlled by only two parameters and, therefore, are arbitrarily interrelated (i.e., only two of these three key moments are free to vary independently). Although the expanded form of the  $S_U$  advanced by Ramírez, Misra, and Field (2003) allows for any mean and variance to be freely associated with the SK values permitted by the original parameterization of this family, it is far from encompassing all theoretically feasible SK combinations.

This research contributes to the yield and price distribution modeling literature through the following measures:

1. Introduce a family of distributions ( $S_B$ ) that perfectly complements the  $S_U$  in its coverage of the SK space, and that spans significant regions of the space not covered by any other distribution that has been used for the modeling and simulation of crop yields.

2. Derive expanded forms of this  $S_B$  family and of the Beta distribution, which are analogous to Ramirez, Misra, and Field's (2003) reparameterization of the  $S_U$ , so that they, too, can allow for any mean and variance to be freely associated with their possible SK values.
3. Assess the importance of MVSK coverage in determining a model's flexibility to adequately represent a wide range of distributional shapes.

### The $S_U$ - $S_B$ System

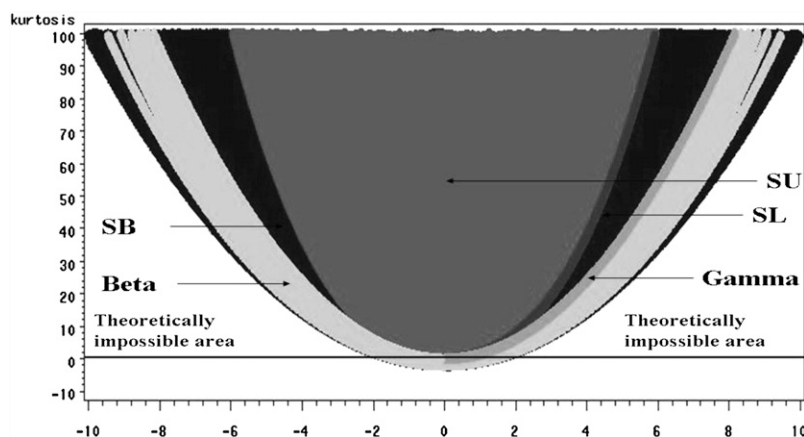
Johnson (1949) introduced the  $S_U$  and the  $S_B$  families of distributions and showed that, together, they span the entire SK space. Figure 1 is constructed on the basis of the formulas for the skewness and kurtosis of the  $S_U$  and  $S_B$  distributions, which were also first derived by Johnson (1949). Specifically, SK pairs are computed and plotted for very fine grids of the parameter spaces corresponding to these two distributions. The same procedure is followed for the Beta, Gamma, and Log-normal ( $S_L$ ). It is observed that the lower bound of the  $S_B$  is at the boundary for the theoretically feasible SK space ( $K = S^2 - 2$ ). This plotting illustrates Johnson's claim of a comprehensive coverage of the theoretically feasible SK space by the  $S_U$ - $S_B$  families.

However, in the parameterizations proposed by Johnson, each of those SK combinations is arbitrarily associated with a fixed set of mean-variance values. Ramirez and McDonald (2006)

outline a reparameterization technique that expands any probability distribution by two parameters that specifically and uniquely control the mean and variance without affecting the range of skewness and kurtosis values that can be accommodated. The expanded distribution obtained through this reparameterization can therefore model any conceivable mean and variance in conjunction with the set of SK combinations allowed by the original distribution. For the purposes of this study, the technique is first applied to the  $S_U$  and  $S_B$  families. This yields a system that can model all theoretically feasible MVSK combinations. The reparameterization begins with the original two-parameter families (Johnson, 1949):

- (1)  $Z = \gamma + \delta \times \sinh^{-1} Y$  for the  $S_U$  distribution
- (2)  $Z = \gamma + \delta \times \ln[Y/(1 - Y)]$  for the  $S_B$  distribution

where  $Y$  is a nonnormally distributed random variable based on a standard normal variable ( $Z$ ). In other words, the original  $S_U$  and  $S_B$  distributions are derived from transformations of a standard normal density. Their pdfs, which are also provided by Johnson (1949), are obtained by substituting  $Z$  in Equation (1) (for the  $S_U$ ) or Equation (2) (for the  $S_B$ ) into a standard normal density and multiplying the resulting equation by the derivative of Equation (1) (for the  $S_U$ ) or Equation (2) (for the  $S_B$ ) with respect to  $Y$ . In the mathematical statistics



**Figure 1.**  $S_U$ ,  $S_L$ ,  $S_B$ , Beta, and Gamma Distributions in the SK Plane; the  $S_B$  Distribution Allows all SK Combinations in the Beta and Gamma Areas as Well

literature, this is commonly known as the transformation technique for deriving pdfs.

Note that from Equations (1) and (2) it follows that:

$$(3) \quad Y = \sinh\left(\frac{Z - \gamma}{\delta}\right) = \sinh(N) \quad \text{for the } S_U \text{ distribution}$$

$$(4) \quad Y = \frac{\exp\left(\frac{Z - \gamma}{\delta}\right)}{\left(1 + \exp\left(\frac{Z - \gamma}{\delta}\right)\right)} = \frac{e^N}{(1 + e^N)} \quad \text{for the } S_B \text{ distribution}$$

where  $N$  is a normal random variable with mean  $-\frac{\gamma}{\delta}$  and variance  $\frac{1}{\delta^2}$ . The above equations express the  $S_U$  and  $S_B$  random variables ( $Y$ ) as a function of a normal, and can be used for simulating draws from their probability distributions. Johnson (1949) also provides the formulas for computing their means and variances, which will be denoted by  $F_{SU}$  and  $F_{SB}$  (for the means) and  $G_{SU}$  and  $G_{SB}$  (for the variances). These formulas are simple but lengthy trigonometric functions of  $\gamma$  and  $\delta$ . The skewness and kurtosis of the  $S_U$  and  $S_B$  distributions are functions of  $\gamma$  and  $\delta$  as well. All formulas and a Gauss program to compute the first four central moments of both distributions for given values for  $\gamma$  and  $\delta$  are available from the authors.

The random variables ( $Y$ ) corresponding to each of the two distributions given in Equations (3) and (4) are then standardized by subtracting their means and dividing by their standard deviations:

$$(5) \quad Y^S = \frac{\sinh(N) - F_{SU}}{G_{SU}^{1/2}} \quad \text{for the } S_U \text{ distribution}$$

$$(6) \quad Y^S = \frac{\frac{e^N}{(1 + e^N)} - F_{SB}}{G_{SB}^{1/2}} \quad \text{for the } S_B \text{ distribution}$$

Note that the standardized  $S_U$  and  $S_B$  variables ( $Y^S$ ) will always have a mean of zero and a variance of one, i.e., the parameters  $\gamma$  and  $\delta$  no longer affect the mean or the variance of their distributions. However, because standardization only involves subtracting from and dividing the original random variables ( $Y$ ) by constants, the distributions corresponding to these standardized variables ( $Y^S$ ) can still accommodate the same sets of skewness-kurtosis combinations

allowed by the original  $S_U$  and  $S_B$  families. The final step in the reparameterization process is to expand the  $Y^S$  distributions so that, instead of being zero and one, their means and variances can be controlled by parameters or by parametric functions of explanatory variables. This is accomplished by multiplying  $Y^S$  times the parameter or parametric function representing the variance and then subtracting the parametric function representing the mean:

$$(7) \quad Y_t^F = \sigma_t Y_t^S - M_t = (Z_t \sigma) Y_t^S - (X_t \beta)$$

where  $Y^S$  is as defined in Equations (5) and (6) for the  $S_U$  and  $S_B$  distributions,  $t = 1, \dots, T$  denotes the observations on the explanatory variables, and  $Y_t^F$  represents the final random variables of interest. From Equation (7), note that for both reparameterized variables:

$$(8) \quad E[Y_t^F] = M_t = X_t \beta \quad \text{and} \quad V[Y_t^F] = \sigma_t^2 = (Z_t \sigma)^2$$

where  $X_t$  and  $Z_t$  represent vectors of explanatory variables believed to affect the means and variances of the distributions, and  $\beta$  and  $\sigma$  are conformable parameter vectors. That is, the mean and variance of the reparameterized  $S_U$  and  $S_B$  random variables ( $Y_t^F$ ) are uniquely controlled by  $M_t$  and  $\sigma_t^2$ , while  $\gamma$  and  $\delta$  determine their skewness and kurtosis according to the formulas provided by Johnson (1949) for the original  $S_U$  and  $S_B$  distributions. Therefore, the reparameterized  $S_U$ - $S_B$  system can accommodate any theoretically possible MVSK combination.

As noted previously, Figure 1 illustrates the SK regions covered by the  $S_U$  and  $S_B$ , as well as three other commonly used distributions. The  $S_L$  or Log-normal distribution, which is also a part of the original Johnson system, only spans the curvilinear boundary between the  $S_U$  and  $S_B$ . The Gamma distribution only spans a curvilinear segment on the upper right quadrant of the SK plane as well. Like the  $S_L$ , the Gamma can be adapted to cover the mirror image of this segment on the upper left quadrant. However, the combinations of SK values allowed by it are still extremely limited.

Although the Beta covers a nonnegligible area of the SK plane, note that the  $S_B$  can accommodate all SK combinations allowed by it.



In fact, the Beta region is quite narrower than the  $S_B$ 's (i.e., the Beta only covers a subset of the SK area spanned by the  $S_B$ ). Thus, in general, one might expect the  $S_B$  to be a more applicable model than the Beta. However, because higher-order moments also affect distributional shape, it is possible that, in some applications, a similarly parameterized Beta would provide for a better model than the  $S_B$ . Another difference that could affect their relative performance in a particular case is the fact that the Beta is a bounded distribution, whereas the  $S_B$  is not.

In short, even if the Beta distribution is reparameterized using the previously discussed technique, because it only spans a relatively small subset of the empirically possible SK space, it may not serve as an acceptable model in some cases. However, it is also possible that it would provide for a better model than the  $S_B$  and the  $S_U$  in other applications. Considering its significant coverage of the SK space and the potential role of the higher-order moments and support characteristics of a distribution on goodness-of-fit, the Beta is selected as the alternative candidate model for the comparative evaluation of the  $S_U$ - $S_B$  system conducted in this study. The expanded parameterization of the Beta distribution is derived in the following section.

### The Expanded Beta Distribution

An expanded parameterization of the Beta distribution that can accommodate any mean and variance in conjunction with all SK combinations allowed by the original Beta is obtained by applying the same technique used for the  $S_U$  and  $S_B$ . In the case of the Beta distribution:

$$(9) \quad \begin{aligned} E[Y] &= \delta / (\delta + \lambda) = F_B \quad \text{and} \\ V[Y] &= \frac{\delta \lambda}{(\delta + \lambda + 1)(\delta + \lambda)^2} = G_B \end{aligned}$$

Thus, the transformation from the original Beta-distributed variable ( $Y$ ) into the random variable exhibiting the expanded Beta distribution ( $Y_t^F$ ) is:

$$(10) \quad Y_t^F = \sigma_t(Y - F_B) / G_B^{1/2} + M_t$$

The pdf for  $Y_t^F$  is obtained through a straightforward application of the transformation technique, which leads to the following log-likelihood function:

$$(11) \quad \begin{aligned} LL_B &= \sum_{t=1}^T \ln \left| \frac{\sqrt{G_B}}{\sigma_t^2} \right| + n \ln \Gamma(\delta + \lambda) \\ &\quad - n \ln \Gamma(\delta) - n \ln \Gamma(\lambda) \\ &\quad + (\delta - 1) \times \left( \sum_{t=1}^T \ln P_t \right) \\ &\quad + (\lambda - 1) \left( \sum_{t=1}^T \ln(1 - P_t) \right), \end{aligned}$$

where  $P_t = \frac{\delta}{\delta + \lambda} + \frac{(Y_t^F - M_t) \times \sqrt{G_B}}{\sigma_t^2}$  and  $\Gamma$  represent the Gamma function. As in the case of the expanded  $S_U$  and  $S_B$ ,  $E[Y_t^F] = M_t$  and  $V[Y_t^F] = \sigma_t^2$  and  $M_t$  and  $\sigma_t^2$  can be specified as linear functions of relevant explanatory variables.

### Estimation of the $S_U$ - $S_B$ System

Estimation of the  $S_U$ - $S_B$  system can also be accomplished by maximum likelihood. Since both originate from normal random variables ( $N$ ), the transformation technique (Mood, Graybill, and Boes 1974) can be applied to derive their probability distribution functions. According to this technique, the pdf of the transformed random variable ( $Y_t^F$ ) is given by:

$$(12) \quad \begin{aligned} P(Y_t^F) &= \left| \frac{\partial(q^{-1}(Y_t^F))}{\partial Y_t^F} \right| \times P(q^{-1}(Y_t^F)) \\ &= J(Y_t^F) \times P(q^{-1}(Y_t^F)), \end{aligned}$$

where  $q^{-1}(Y_t^F)$  is the inverse of the transformation of  $N$  into  $Y_t^F$  (i.e., the function relating  $N$  to  $Y_t^F$ ,  $P(q^{-1}(Y_t^F))$  is the pdf of an independently and identically distributed normal random variable  $N$  with mean  $(-\frac{\gamma}{\delta})$  and variance  $\delta^{-2}$  evaluated at  $q^{-1}(Y_t^F)$ , and  $J(Y_t^F)$  is the Jacobian of the transformation—that is, the derivative of  $q^{-1}(Y_t^F)$  with respect to  $Y_t^F$ . Specifically, for the  $S_U$ ,  $N = q_{SU}^{-1}(Y_t^F)$  is found by substituting Equation (5) into Equation (7) and solving for  $N$ :

$$(13) \quad \begin{aligned} N &= q_{SU}^{-1}(Y_t^F) = \sinh^{-1} \{R_{SUt}\} \quad \text{where} \\ R_{SUt} &= \frac{(Y_t^F - X_t\beta) \times G_{SU}^{1/2}}{Z_t\sigma} + F_{SU}, \end{aligned}$$

and  $F_{SU}$  and  $G_{SU}$  are as defined previously. The Jacobian is obtained by taking the absolute value of the derivative of Equation (13) with respect to  $Y_t^F$ , which yields:

$$(14) \quad J_{SU}(Y_t^F) = \frac{G_{SU}^{1/2}}{Z_t \sigma (1 + R_{SUt}^2)^{1/2}}.$$

The  $S_U$  pdf is then obtained by substituting Equation (13) into a normal density with mean  $(-\frac{\gamma}{\delta})$  and variance  $\delta^{-2}$  and premultiplying the result by Equation (14).

Analogously, for the  $S_B$ , Equation (6) is substituted into Equation (7) to obtain:

$$(15) \quad N = q_{SB}^{-1}(Y_t^F) = \ln \left\{ \frac{R_{SBt}}{1 - R_{SBt}} \right\} \quad \text{where}$$

$$R_{SBt} = \frac{(Y_t^F - X_t \beta)}{Z_t \sigma} \times \frac{G_{SB}^{1/2}}{Z_t \sigma} + F_{SB},$$

and  $F_{SB}$  and  $G_{SB}$  are as defined previously. The Jacobian is:

$$(16) \quad J_{SB}(Y_t^F) = \frac{G_{SB}^{1/2}}{Z_t \sigma R_{SBt} (1 - R_{SBt})}.$$

Substituting Equation (15) into a normal density with mean  $(-\frac{\gamma}{\delta})$  and variance  $\delta^{-2}$  and premultiplying the result by Equation (16) renders the  $S_B$  pdf.

The log-likelihood functions are found by taking the natural logarithms of the resulting pdfs and adding over the  $T$  observations:

$$(17) \quad \sum_{t=1}^T \ln \{P(Y_t^F)\} = 0.5 \sum_{t=1}^T \ln(G_t) - 0.5 \delta^2 \sum_{t=1}^T H_t^2, \quad \text{where:}$$

$$(18) \quad G_t = \frac{\delta^2 G_{SU}}{2\pi(Z_t \sigma)^2 (1 + R_{SUt}^2)} \quad \text{and}$$

$$H_t = \ln \left[ R_{SUt} + \sqrt{1 + R_{SUt}^2} \right] + \frac{\gamma}{\delta}$$

$$= \sinh^{-1}(R_{SUt}) + \frac{\gamma}{\delta} \quad \text{for the } S_U \text{ and}$$

$$(19) \quad G_t = \frac{\delta^2 G_{SB}}{2\pi(Z_t \sigma)^2 R_{SBt}^2 (1 - R_{SBt})^2} \quad \text{and}$$

$$H_t = \ln[R_{SBt} / (1 - R_{SBt})] + \frac{\gamma}{\delta} \quad \text{for the } S_B,$$

respectively;  $G_t > 0$ ; and  $G_{SU}$ ,  $R_{SUt}$ ,  $G_{SB}$ , and  $R_{SBt}$  are as defined previously.

An adjustment that facilitates estimation and interpretation is redefining the distributional shape parameters as follows: for the  $S_U$   $\gamma = -\mu$ , for the  $S_B$   $\gamma = \mu$ , and for both families  $\delta = 1/\theta$ . Then, in all cases  $\mu < 0$ ,  $\mu = 0$ , and  $\mu > 0$  are associated with negative, zero,

and positive skewness, and both families approach a normal distribution as  $\theta$  goes to zero. This also allows for testing the null hypothesis of normality as  $H_0: \theta = \mu = 0$ . In addition, for the purposes of estimation, the following parameter range restrictions are recommended based on the authors' experience:  $0 < \theta < 1.5$  and  $-10 < \mu < 10$  for the  $S_U$ ;  $0 < \theta < 100$  and  $-7.5 < \mu < 7.5$  for the  $S_B$ . Finally, it is noted that the Gauss programs used for Maximum Likelihood estimation of the  $S_U$  and the  $S_B$  distributions, for computing their mean, variance skewness, and kurtosis given estimated or assumed parameter values, and for simulating random draws from them are available from the authors upon request. These programs can be easily translated into MatLab or SAS-IML.

## Validating the Proposed Theoretical Framework

Ramirez and McDonald (2006) suggest that MVSK space coverage is key to a model's flexibility to adequately represent a wide range of distributions. This section empirically evaluates that hypothesis. Specifically, it explores to what extent different distributions with similar MVSK coverage can "substitute" for each other in practice. This is accomplished through large sample comparisons between the  $S_U$ ,  $S_B$ , and Beta, which are the three distributions that provide for substantial coverage of the MVSK space (Figure 1).

Parametric models of farm-level yields based on these three families of distributions are estimated using Gauss programs. The data, obtained from the University of Illinois Endowment Farms database, includes 26 corn farms located in twelve counties across that state. Data are available from 1959 to 2003, with sample sizes ranging from 20 to 45 observations. The mean and standard deviations are specified as second- and first-degree polynomial functions of time:

$$(20) \quad M_t = X_t \beta = \beta_0 + \beta_1 t + \beta_2 t^2 \quad \text{and} \quad \sigma_t = (Z_t \sigma)$$

$$= \sigma_0 + \sigma_1 t; \quad t = 1, \dots, T$$

All nonnormal models initially include seven parameters ( $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ ,  $\sigma_0$ ,  $\sigma_1$ ,  $\theta$ , and  $\mu$  in the case of the  $S_U$ - $S_B$  system and  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ ,  $\sigma_0$ ,  $\sigma_1$ ,

$\delta$ , and  $\lambda$  in the case of the Beta). Normal models with the same mean and standard deviation specifications are also estimated and included in the comparison. Select statistics about those models are presented in Table 1.

The null hypothesis of yield normality is tested through likelihood ratio tests of the normal model versus the nonnormal model with the highest maximum log-likelihood function value (MLLFV). This is possible because the Normal model is nested to all three nonnormal models. If this hypothesis cannot be rejected ( $\alpha = 0.10$ ), it is assumed that the underlying pdf is Normal. Otherwise, it is assumed that the true distribution is the nonnormal with the highest MLLFV.

Although this is not a formal statistical test for the superiority of a nonnormal model over another, it has been routinely used to rank fit (Norwood, Roberts, and Lusk, 2004). Following this criterion, five of the 26 underlying distributions are categorized as Normal, six as  $S_U$ , seven as  $S_B$ , and eight as Beta (Table 1).

For the purpose of the analysis, the corresponding  $S_U$ ,  $S_B$ , and Beta models are assumed to be the true distributions and used to generate the data for the next phase. Specifically, 21 datasets of 100,000 observations each are simulated on the basis of the six  $S_U$ , seven  $S_B$ , and eight estimated Beta models. The fact that the distributional shapes used in this evaluation are

**Table 1.** Select Statistics for Illinois Farm-Level Corn Yield Models Based on  $S_U$ ,  $S_B$ , Beta, and Normal Distributions

Farm Label	Sample Size					Likelihood	Final Model
		$S_U$ MLLFV	$S_B$ MLLFV	Beta MLLFV	Normal MLLFV	Ratio Test Statistic <sup>a</sup>	
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>	N
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>	N
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>	N
M	37	-169.08	-169.00	-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	-150.09	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	-127.78	-126.34	-125.69	-131.64	11.91 <sup>3</sup>	Beta
W	29	-131.22	-131.24	-131.20	-132.56	2.71 <sup>0</sup>	N
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>	N

Note: MLLFV refers to the maximum log-likelihood function value.

<sup>a</sup> The likelihood ratio test statistic compares the nonnormal model with the highest MLLFV with the normal model. The superscripts 1, 2, and 3 denote rejection of the null hypothesis of normality at the 10%, 5%, and 1% levels, respectively, according to the likelihood ratio test, while the superscript 0 indicates nonrejection at the 10% level. If the null hypothesis of normality is rejected at the 10% level, the final model is the one with the highest MLLFV, otherwise the final model is the normal.



empirically motivated (i.e., derived from parametric models that have been estimated on the basis of actual yield data) enhances the credibility of the analysis.

The following simulation formulas are based on Equations (5), (6), and (7) for the proposed system and Equation (10) for the Beta distribution:

$$(21) \quad SS_U = M_t + \left\{ \sigma_t [\sinh(\theta[Z + \mu]) - F_{SU}] \div \theta \sqrt{G_{SU}} \right\}$$

$$(22) \quad SS_{SB} = M_t + \left\{ \sigma_t \exp(\theta[Z - \mu]) \div \sqrt{G_{SB}} [1 + \exp(\theta[Z - \mu]) - F_{SB}] \right\}$$

$$(23) \quad S_B = M_t + \sigma_t \left\{ (B - F_B) \div \sqrt{G_B} \right\}$$

where  $Z$  is a draw from a standard Normal and  $B$  is a draw from a Beta distribution with parameters  $\delta$  and  $\lambda$ .

Next, a second round of  $S_U$ ,  $S_B$ , Beta, and Normal models are estimated on the basis of each of the 21 simulated datasets. Key statistics about those models are presented in Table 2 (data-generating process =  $S_B$ ), Table 3 (data-generating process = Beta), and Table 4 (data-generating process =  $S_U$ ). As expected, in each of the 21 cases, the model with the highest MLLFV is the one based on the probability distribution used to simulate the data.

In the case of the seven sets of models corresponding to the  $S_B$ -generated datasets (Table 2), the MLLFVs of the Beta models are relatively close to those of the  $S_B$  models, with the differences averaging 1.02 units. At 2.53 units, the average MLLFV difference between the  $S_B$  and  $S_U$  models is considerably larger. The Normal models show substantially lower MLLFVs than any of the three nonnormal models in all cases.

In the case of the eight sets of models corresponding to the Beta-generated datasets (Table 3), with differences averaging 0.39 units, the MLLFVs of the  $S_B$  models are very close to those of the Beta models. At 1.26 units, the average MLLFV difference between the Beta and the  $S_U$  models is over four times larger. As before, the Normal models show much lower MLLFVs than any of the three nonnormal models.

In the case of the six sets of models corresponding to the  $S_U$ -generated datasets (Table 4), both the  $S_B$  and the Beta models yield MLLFVs that are substantially lower than those of the  $S_U$  models. On average, the MLLFVs are 11.71 units lower in the  $S_B$ , 12.60 units lower in the Beta, and 13.94 units lower in the Normal models.

In short, the MLLFV comparisons suggest that the  $S_U$  model is not a close substitute for either the  $S_B$  or the Beta, and that the  $S_B$  and the Beta models are poor surrogates for the  $S_U$ . In contrast, it appears that the  $S_B$  and the Beta models could be acceptable substitutes for each other, with the  $S_B$  being a better surrogate for the Beta than the Beta is for the  $S_B$ . These findings support the hypothesis that distributions with similar MVSK coverage are better able to “substitute” for each other. However, the question remains as to how well these nonnormal models can substitute for each other.

To answer this question, the cumulative distribution functions (cdfs) implied by the second-round  $S_U$ ,  $S_B$ , Beta, and Normal models are obtained for each of the 21 cases, also through simulation. The “true” cdfs are also plotted using the correct distribution and the exact parameters underlying each of the 21 data-generating processes. Two main statistics related to those cdfs are also presented in Tables 2, 3, and 4. AD is the average of 125 vertical percentage distances between the true and the estimated cdf. Distances are computed for yield values ranging from 25% to 150% of the average yields at equal 1% intervals (cdf values outside of that range are negligible in all cases). MD represents the maximum of those 125 vertical distances. Cdf values are expressed to range from zero to 100%, instead of zero to one.

Table 2 contains the statistics for the seven cases when the underlying data-generating process is  $S_B$ . As expected, the estimated  $S_B$  cdfs are very close to the cdfs obtained on the basis of the true models (AD of 0.03% and MD of 0.17%, on average). When  $S_U$  models are used to approximate the  $S_B$ , across the seven cases the AD averages 1.62% and the MD averages 4.50%. As anticipated, because the SK

**Table 2.** Key Statistics about the  $S_U$ ,  $S_B$ , Beta, and Normal Models Estimated on the Basis of the Seven  $S_B$ -Simulated Datasets

Model	FARM B (DGP = $S_B$ )				FARM C (DGP = $S_B$ )				FARM J (DGP = $S_B$ )			
	$S_U$	$S_B$	Beta	Normal	$S_U$	$S_B$	Beta	Normal	$S_U$	$S_B$	Beta	Normal
MLLFV	-156.70	-156.69	-158.02	-172.58	-168.81	-163.41	-164.79	-173.64	-181.44	-181.44	-181.97	-190.54
Skewness	-3.25	-2.77	-1.46	0	-3.24	-0.63	-0.71	0	-2.01	-1.93	-1.19	0
Kurtosis	23.27	14.25	3.04	0	23.15	-0.74	-0.30	0	7.92	7.04	2.08	0
AD	0.07%	0.02%	0.95%	3.54%	2.65%	0.02%	1.13%	2.67%	0.04%	0.01%	0.78%	3.44%
MD	0.31%	0.22%	4.58%	14.28%	7.67%	0.25%	4.83%	10.21%	0.14%	0.07%	2.77%	10.78%

Model	FARM K (DGP = $S_B$ )				FARM N (DGP = $S_B$ )				FARM O (DGP = $S_B$ )			
	$S_U$	$S_B$	Beta	Normal	$S_U$	$S_B$	Beta	Normal	$S_U$	$S_B$	Beta	Normal
MLLFV	-180.46	-176.62	-178.09	-192.06	-177.98	-171.56	-173.61	-178.09	-180.70	-178.92	-179.29	-185.33
Skewness	-7.42	-1.05	-1.50	0	0.36	-0.09	-0.16	0	-2.13	-0.81	-0.75	0
Kurtosis	176.36	0.17	2.62	0	0.23	-1.21	-1.09	0	9.02	-0.06	0.01	0
AD	4.20%	0.04%	1.36%	5.04%	2.19%	0.04%	0.74%	2.18%	1.60%	0.02%	0.53%	2.94%
MD	11.26%	0.29%	3.80%	14.06%	6.41%	0.14%	2.51%	6.70%	4.23%	0.13%	1.82%	8.92%

Model	FARM P (DGP = $S_B$ )				AVERAGES (DGP = $S_B$ )			
	$S_U$	$S_B$	Beta	Normal	$S_U$	$S_B$	Beta	Normal
MLLFV	-185.85	-185.56	-185.59	-187.83	-175.99	-173.46	-174.48	-182.87
Skewness	-0.96	-0.66	-0.67	0	-2.66	-1.13	-0.92	0.00
Kurtosis	1.70	0.19	0.32	0	34.52	2.81	0.95	0.00
AD	0.62%	0.04%	0.21%	2.13%	1.62%	0.03%	0.81%	3.13%
MD	1.49%	0.07%	0.57%	5.56%	4.50%	0.17%	2.98%	10.07%

Notes: DGP is the data-generating process; MLLFV refers to the maximum log-likelihood function value; AD is the average of 125 vertical percentage distances between the true and the estimated cdfs; and MD represents the maximum of those 125 vertical distances. Distances are computed for yield values ranging from 25% to 150% of the mean yields at equal 1% intervals. Cumulative yield probability values outside of that range are negligible in all cases.

**Table 3.** Key Statistics about the  $S_U$ ,  $S_B$ , Beta, and Normal Models Estimated on the Basis of the Eight Beta-Simulated Datasets

Model	FARM E (DGP = Beta)				FARM G (DGP = Beta)				FARM M (DGP = Beta)			
	$S_U$	$S_B$	Beta	Normal	$S_U$	$S_B$	Beta	Normal	$S_U$	$S_B$	Beta	Normal
MLLFV	-181.38	-181.00	-180.71	-190.63	-172.60	-172.16	-171.79	-182.69	-183.60	-183.20	-183.14	-186.53
Skewness	-2.78	-1.71	-1.56	0	-3.16	-1.83	-1.59	0	-1.24	-0.84	-0.81	0
Kurtosis	16.32	3.94	3.56	0	21.83	4.55	3.62	0	2.86	0.51	0.54	0
AD	0.83%	0.49%	0.05%	3.44%	0.86%	0.49%	0.00%	2.91%	0.58%	0.16%	0.02%	1.93%
MD	2.88%	1.79%	0.24%	10.73%	3.46%	2.11%	0.04%	11.22%	1.75%	0.62%	0.08%	6.36%

Model	FARM Q (DGP = Beta)				FARM T (DGP = Beta)				FARM U (DGP = Beta)			
	$S_U$	$S_B$	Beta	Normal	$S_U$	$S_B$	Beta	Normal	$S_U$	$S_B$	Beta	Normal
MLLFV	-173.72	-173.09	-172.41	-186.76	-178.86	-178.21	-177.71	-191.80	-183.79	-181.38	-181.12	-186.05
Skewness	-3.98	-2.09	-1.80	0	4.45	-2.00	-1.79	0	-1.41	-0.61	-0.55	0
Kurtosis	37.64	5.75	4.64	0	49.11	5.09	4.66	0	3.72	-0.50	-0.53	0
AD	0.98%	0.71%	0.01%	3.46%	1.52%	0.97%	0.13%	4.56%	2.02%	0.51%	0.01%	2.83%
MD	3.85%	2.97%	0.04%	14.16%	4.99%	3.15%	0.55%	13.74%	4.34%	1.23%	0.05%	6.89%

Model	FARM V (DGP = Beta)				FARM Y (DGP = Beta)				AVERAGES (DGP = Beta)			
	$S_U$	$S_B$	Beta	Normal	$S_U$	$S_B$	Beta	Normal	$S_U$	$S_B$	Beta	Normal
MLLFV	-173.22	-172.52	-172.02	-186.98	-186.69	-184.53	-184.06	-191.69	-179.23	-178.26	-177.87	-187.89
Skewness	-4.74	-2.05	-1.85	0	-2.41	-0.95	-0.84	0	-1.91	-1.51	-1.35	0.00
Kurtosis	57.24	5.30	5.00	0	11.79	0.22	0.07	0	25.06	3.11	2.70	0.00
AD	1.11%	0.66%	0.03%	3.09%	1.76%	0.58%	0.05%	2.80%	1.21%	0.57%	0.04%	3.13%
MD	5.01%	3.23%	0.18%	14.11%	5.00%	1.95%	0.30%	9.14%	3.91%	2.13%	0.19%	10.79%

Notes: DGP is the data-generating process; MLLFV refers to the maximum log-likelihood function value; AD is the average of 125 vertical percentage distances between the true and the estimated cdfs; and MD represents the maximum of those 125 vertical distances. Distances are computed for yield values ranging from 25% to 150% of the mean yields at equal 1% intervals. Cumulative yield probability values outside of that range are negligible in all cases.

**Table 4.** Key Statistics about the  $S_U$ ,  $S_B$ , Beta, and Normal Models Estimated on the Basis of the Six  $S_U$ -Simulated Datasets

Model	FARM A (DGP = $S_U$ )				FARM D (DGP = $S_U$ )				FARM I (DGP = $S_U$ )			
	$S_U$	$S_B$	Beta	Normal	$S_U$	$S_B$	Beta	Normal	$S_U$	$S_B$	Beta	Normal
MLLFV	-165.79	-174.33	-176.38	-177.93	-175.40	-176.05	-176.76	-178.60	-167.28	-185.23	-185.23	-185.23
Skewness	-3.94	-0.31	-0.21	0	-1.18	-0.58	-0.35	0	-0.83	0	0	0
Kurtosis	58.33	0.14	0.03	0	3.44	0.60	0.15	0	369.63	0	0	0
AD	0.06%	2.96%	3.74%	3.91%	0.02%	0.95%	1.42%	2.03%	0.08%	5.04%	5.04%	5.04%
MD	0.25%	11.34%	12.92%	14.02%	0.09%	3.11%	4.61%	6.73%	0.18%	15.60%	15.60%	15.60%
Model	FARM R (DGP = $S_U$ )				FARM S (DGP = $S_U$ )				FARM X (DGP = $S_U$ )			
	$S_U$	$S_B$	Beta	Normal	$S_U$	$S_B$	Beta	Normal	$S_U$	$S_B$	Beta	Normal
MLLFV	-176.92	-178.85	-180.59	-184.67	-181.43	-183.89	-184.68	-185.28	-186.37	-225.13	-225.13	-225.13
Skewness	-2.11	-0.73	-0.45	0	-1.39	-0.25	-0.13	0	-29.46	0	0	0
Kurtosis	10.14	0.95	0.27	0	5.71	0.09	-0.01	0	10369.3	0	0	0
AD	0.10%	1.86%	2.57%	3.80%	0.06%	1.56%	1.93%	2.45%	0.04%	9.25%	9.25%	9.25%
MD	0.35%	5.75%	7.56%	10.69%	0.17%	5.42%	6.47%	7.68%	0.18%	24.85%	24.85%	24.85%
AVERAGES (DGP = $S_U$ )												
Model	$S_U$	$S_B$	Beta	Normal								
MLLFV	-175.53	-187.24	-188.13	-189.47								
Skewness	-6.49	-0.31	-0.19	0.00								
Kurtosis	1802.76	0.30	0.07	0.00								
AD	0.06%	3.60%	3.99%	4.41%								
MD	0.20%	11.01%	12.00%	13.26%								

Notes: DGP is the data-generating process; MLLFV refers to the maximum log-likelihood function value; AD is the average of 125 vertical percentage distances between the true and the estimated cdfs; and MD represents the maximum of those 125 vertical distances. Distances are computed for yield values ranging from 25% to 150% of the mean yields at equal 1% intervals. Cumulative yield probability values outside of that range are negligible in all cases.

regions spanned by the  $S_U$  and the  $S_B$  do not overlap (Figure 1), in general, the  $S_U$  is a poor surrogate for the  $S_B$ . However, the  $S_U$  approximation of the  $S_B$  is much better in two of the cases when the SK combination is near the  $S_U$ - $S_B$  boundary (Figure 2).

Alternatively, when Beta models are used to approximate the  $S_B$ , AD averages 0.81% and MD averages 2.98%. As anticipated, because the Beta spans only part of the SK region covered by the  $S_B$  (Figure 1) and two of the  $S_B$  data-generating processes exhibit SK values that cannot be accommodated by the Beta (Figure 2), this distribution does not generally provide a good approximation of the  $S_B$ . On average, however, the Beta is a substantially better surrogate for the  $S_B$  than the  $S_U$ . With average ADs of 3.13% and MDs of 10.07%, the normal cdfs are by far the worst models for an underlying  $S_B$  process.

Table 3 presents statistics for the eight instances when the data-generating process is Beta. The estimated Beta cdfs are again very close to the cdfs obtained on the basis of the true models (AD of 0.04% and MD of 0.19%, on average). When  $S_U$  models are used to approximate the Beta, the AD averages 1.21% and the MD averages 3.91%. Because the SK regions spanned by the Beta and the  $S_U$  do not overlap and are in fact separated by a widening segment of the SK space (Figure 1), the  $S_U$  is consistently a poor surrogate for the Beta.

In contrast, when  $S_B$  models are used to approximate the Beta, AD averages 0.57% and MD averages 2.13%. As anticipated, because the  $S_B$  spans all of the SK area covered by the Beta (Figure 1), it provides for a relatively satisfactory approximation of that distribution. Figure 3 provides a visual cue of the closeness with which the typical  $S_B$  model can replicate a true Beta cdf. All vertical differences in the lower one-third of the cdf are in fact less than 1.1%. That is, the  $S_B$  model can predict cumulative probability at any point within the lower third of the true cdf with a margin of error of 1.1% or less. This is particularly noteworthy because in most cases the lower (left) tail is the relevant segment of the cdf for the purposes of risk analysis. With average ADs of 3.13% and MDs of 10.79%, the normal cdfs are by far the worst models for an underlying Beta process as well.

The statistics for the six cases when the data-generating process is  $S_U$  are shown in Table 4. The estimated  $S_U$  cdfs are again very close to the cdfs obtained on the basis of the true models (AD of 0.06% and MD of 0.20%, on average). In addition, as expected from the fact that the  $S_B$  and the Beta do not span any of the SK space covered by the  $S_U$ , with ADs averaging 3.60% and 3.99%, respectively, and MDs averaging 11.01% and 12.00%, respectively, they are not good surrogates for the  $S_U$ . Particularly large cdf approximation errors

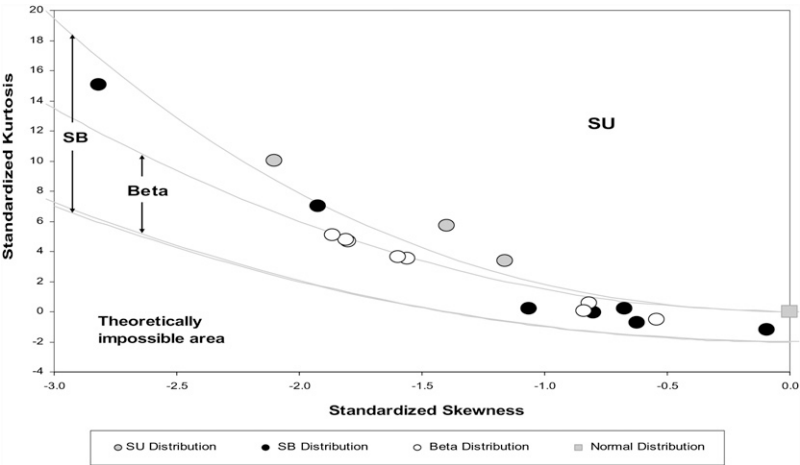
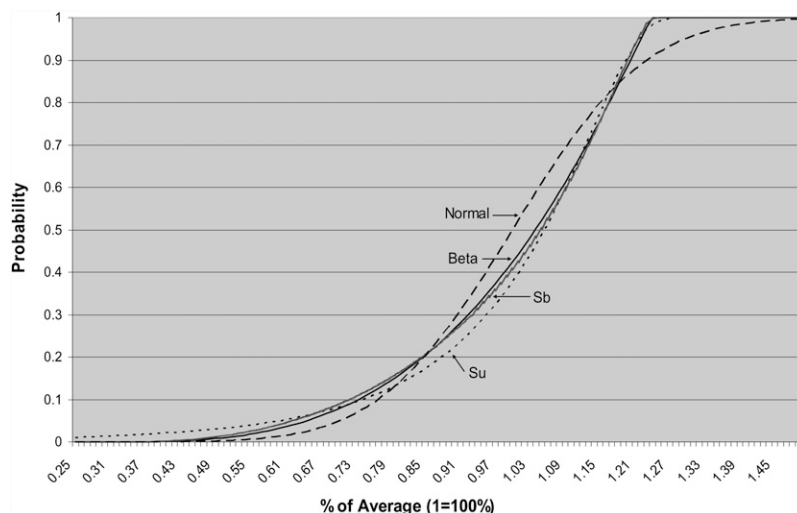


Figure 2. Skewness-Kurtosis Combinations of Estimated Nonnormal Models





**Figure 3.** Estimated  $S_U$ ,  $S_B$ , and Normal versus True (Beta) cdf

are observed in the two cases when the  $S_U$  process is characterized by a very high kurtosis-to-skewness ratio, which are not shown in Figure 2 due to scale limitations. The normal cdfs are again the worst models.

In short, the results presented in Tables 2, 3, and 4 support the working hypothesis that a distribution's capacity to provide for an adequate approximation of another is closely related to its ability to accommodate the underlying MVSK value. These results also suggest that researchers can achieve a relatively small specification error by using parametric models based on distributions which, as a whole, span the entire MVSK space. Specifically, given that there are no other distributions spanning the middle-upper regions of this space (the green area in Figure 1) and that several empirically occurring yield distributions seem to exhibit SK values well into this area, the  $S_U$  should always be considered as a candidate model. In addition, the results indicate that the  $S_B$  is a generally better alternative than the Beta for underlying distributions with SK values on the surrounding regions of the SK space (the blue, yellow, pink, and red areas in Figure 1) because the Beta can only partially cover these remaining regions.

However, if those SK values are in the Beta (i.e., yellow) area and the higher-order moments and support characteristics of the true

underlying distribution are more consistent with the Beta's, a model based on this density would be expected to provide for a somewhat better fit than the  $S_B$ . Likewise, under analogous conditions, the Gamma, Log-normal, and other distributions could improve fit in relation to the  $S_B$ . However, these expected gains are only assured to materialize under large sample conditions. Therefore, when working with small samples, it might be best to simply consider the two most general alternatives (i.e., the  $S_U$  and the  $S_B$ ) as candidate models.

### Economic Relevance

A final issue of interest is the economic relevance of using a more suitable probability distribution model for risk management decisions. This issue is explored on the basis of the results from the previous section. Specifically, the cdfs implied by the second-round  $S_U$ ,  $S_B$ , Beta, and Normal models are used to compute the values, expressed as percentages of the distributions' means, that correspond to the 5, 10, and 20 cdf percentiles (Table 5). In the case of farm E, for example, the Normal,  $S_U$ ,  $S_B$ , and Beta cdfs reach their 5<sup>th</sup> percentile at 72%, 63%, 64%, and 66% of their respective means.

The economic relevance of this information can be related to crop insurance. Assume that a farm manager wants to know the coverage

Table 5. Coverage Levels Implied by True Distribution versus Three Alternatives for Three Different Risk Tolerance Levels

TD/Farm	Risk Tolerance	% Coverage			TD/Farm	Risk Tolerance	% Coverage			S <sub>U</sub>	% Coverage			S <sub>B</sub>	% Coverage			Beta
		Normal	S <sub>U</sub>	S <sub>B</sub>			Normal	S <sub>U</sub>	S <sub>B</sub>		Normal	S <sub>U</sub>	S <sub>B</sub>		Normal	S <sub>U</sub>	S <sub>B</sub>	
Beta/E	5%	0.72	0.63	0.64	Sb/J	5%	0.72	0.67	0.72	0.67	0.72	0.67	0.67	0.67	0.72	0.67	0.67	0.70
	10%	0.78	0.77	0.76		10%	0.78	0.77	0.78	0.79	0.78	0.79	0.79	0.79	0.78	0.79	0.79	0.79
	20%	0.85	0.90	0.88		20%	0.85	0.90	0.86	0.89	0.86	0.89	0.89	0.89	0.86	0.89	0.89	0.88
Beta/G	5%	0.77	0.70	0.70	Sb/K	5%	0.72	0.72	0.63	0.35	0.63	0.35	0.52	0.52	0.63	0.35	0.52	0.55
	10%	0.82	0.82	0.81		10%	0.81	0.81	0.71	0.65	0.71	0.65	0.64	0.64	0.71	0.65	0.64	0.69
	20%	0.88	0.92	0.91		20%	0.90	0.90	0.81	0.89	0.81	0.89	0.81	0.81	0.81	0.89	0.81	0.84
Beta/M	5%	0.73	0.68	0.69	Sb/N	5%	0.69	0.69	0.72	0.71	0.72	0.71	0.73	0.73	0.72	0.71	0.73	0.73
	10%	0.79	0.78	0.77		10%	0.77	0.77	0.79	0.78	0.79	0.78	0.76	0.76	0.79	0.78	0.76	0.77
	20%	0.87	0.88	0.87		20%	0.87	0.88	0.86	0.86	0.86	0.86	0.82	0.82	0.86	0.86	0.82	0.84
Beta/Q	5%	0.76	0.69	0.67	Sb/O	5%	0.70	0.70	0.69	0.59	0.69	0.59	0.62	0.62	0.69	0.59	0.62	0.64
	10%	0.81	0.81	0.79		10%	0.80	0.80	0.76	0.73	0.76	0.73	0.71	0.71	0.76	0.73	0.71	0.73
	20%	0.87	0.93	0.91		20%	0.90	0.90	0.84	0.87	0.84	0.87	0.83	0.83	0.84	0.87	0.83	0.84
Beta/T	5%	0.67	0.54	0.54	Sb/P	5%	0.60	0.60	0.65	0.60	0.65	0.60	0.60	0.60	0.65	0.60	0.60	0.61
	10%	0.74	0.73	0.71		10%	0.73	0.73	0.73	0.72	0.73	0.72	0.70	0.70	0.73	0.72	0.70	0.71
	20%	0.83	0.89	0.87		20%	0.87	0.89	0.82	0.84	0.82	0.84	0.82	0.82	0.82	0.84	0.82	0.83
Beta/U	5%	0.59	0.49	0.52	Su/A	5%	0.53	0.53	0.73	0.73	0.73	0.73	0.74	0.74	0.73	0.73	0.74	0.72
	10%	0.68	0.65	0.63		10%	0.64	0.65	0.79	0.84	0.79	0.84	0.80	0.80	0.79	0.84	0.80	0.78
	20%	0.79	0.82	0.77		20%	0.77	0.82	0.86	0.93	0.86	0.93	0.88	0.88	0.86	0.93	0.88	0.86
Beta/V	5%	0.79	0.71	0.71	Su/D	5%	0.74	0.74	0.74	0.71	0.74	0.71	0.71	0.71	0.74	0.71	0.71	0.72
	10%	0.83	0.83	0.82		10%	0.83	0.83	0.79	0.80	0.79	0.80	0.79	0.79	0.79	0.80	0.79	0.79
	20%	0.89	0.94	0.92		20%	0.92	0.94	0.86	0.89	0.86	0.89	0.87	0.87	0.86	0.89	0.87	0.87
Beta/Y	5%	0.70	0.60	0.62	Su/I	5%	0.65	0.60	0.68	0.77	0.68	0.77	0.68	0.68	0.68	0.77	0.68	0.68
	10%	0.77	0.74	0.72		10%	0.73	0.74	0.75	0.86	0.75	0.86	0.75	0.75	0.75	0.86	0.75	0.75
	20%	0.85	0.88	0.85		20%	0.85	0.88	0.84	0.93	0.84	0.93	0.84	0.84	0.84	0.93	0.84	0.84
Sb/B	5%	0.78	0.75	0.75	Su/R	5%	0.78	0.75	0.67	0.64	0.67	0.64	0.67	0.67	0.67	0.64	0.67	0.67
	10%	0.83	0.84	0.84		10%	0.85	0.84	0.75	0.76	0.75	0.76	0.76	0.76	0.75	0.76	0.76	0.75
	20%	0.89	0.93	0.93		20%	0.92	0.93	0.83	0.88	0.83	0.88	0.86	0.86	0.83	0.88	0.86	0.85
Sb/C	5%	0.74	0.61	0.70	Su/S	5%	0.72	0.72	0.71	0.68	0.71	0.68	0.72	0.72	0.71	0.68	0.72	0.72
	10%	0.80	0.76	0.75		10%	0.79	0.76	0.77	0.78	0.77	0.78	0.79	0.79	0.77	0.78	0.79	0.78
	20%	0.87	0.90	0.85		20%	0.87	0.90	0.85	0.88	0.85	0.88	0.87	0.87	0.85	0.88	0.87	0.86

Table 5. Continued

TD/Farm	Risk Tolerance	% Coverage Normal	% Coverage Su	% Coverage Sb	% Coverage Beta	Average Differences <sup>a</sup>	Normal	Su	Sb	Beta
Su/X	5%	0.61	0.76	0.61	0.61	Beta	0.031	0.021	0.012	0.000
	10%	0.70	0.88	0.70	0.70	Sb	0.036	0.030	0.000	0.019
	20%	0.80	0.97	0.80	0.80	Su	0.063	0.000	0.057	0.061

Notes: TD/Farm is the true underlying distribution and the farm associated with it. % Coverage refers to the coverage level that would have to be selected to achieve a particular level of risk protection (i.e., tolerance) according to each of the four distributions being evaluated.

<sup>a</sup> Averages of the absolute differences between the % coverage levels implied by the true distributions (in the three rows) and the four alternatives (in the four columns).

level that he/she would need to purchase in order to be protected from yield losses that are likely to occur every 20, 10, and 5 years (i.e., for three different levels of risk tolerance: 5%, 10%, and 20%). Although actual coverage levels are limited to 65%, 70%, 75%, 80%, and 85% of the “average proven yield” (i.e., the estimated mean of the yield distribution), exact levels are computed and used for the purposes of this evaluation.

In the case of farm E, for example, the true underlying distribution is Beta and, therefore, the farm manager should choose 66%, 77%, or 88% coverage levels depending on his/her risk tolerance. If this decision was being made on the basis of an estimated  $S_B$  distribution, the selected levels would be 64%, 76%, and 88% (Table 5). In general, in the eight instances when the true underlying distribution is Beta, the coverage levels suggested by the  $S_B$  model are fairly close to the correct ones. The average of the absolute differences between the correct levels and those implied by the  $S_B$  distribution across the 24 cases (eight farms and three risk-tolerance levels) is only 1.2% (Table 5). Given the actual coverage level choices (0.65%, 0.70%, 0.75%, 0.80%, and 0.85%), differences of this magnitude are unlikely to cause an incorrect selection in most cases and, therefore, could be considered relatively unimportant from an economic standpoint.

In the case of farm C, the true underlying distribution is  $S_B$  and, therefore, the farm manager should choose the 70%, 75%, or 85% coverage level depending on his/her risk tolerance. If this decision was made on the basis of an estimated Beta model, the selected levels would be 72%, 79%, and 87%. In general, in the seven instances when the true underlying distribution is  $S_B$ , the coverage levels implied by the Beta are somewhat different from the correct ones (Table 5). The average absolute difference in this case is 1.9%. Differences of this magnitude are more likely to cause incorrect coverage selection in some cases and could therefore be considered somewhat important from an economic standpoint.

Finally, in the six instances when the true underlying distribution is  $S_U$ , the coverage levels implied by the Beta or the  $S_B$  model are

generally quite different from the correct ones (Table 5). In the case of the Beta, the average of the absolute differences is 6.1%. Differences of this magnitude are likely to cause major errors in coverage selection in most cases and, therefore, are economically important.

In short, the conclusions from the economic relevance evaluation are consistent with those of the previous section. If the true distribution underlying the yield data are Beta and management decisions are made on the basis of an estimated  $S_B$  model, the degree of error and its economic implications are relatively minor. If the true distribution is  $S_B$  and decisions are made using a Beta, the errors are somewhat higher and more likely to be economically significant in some cases. Finally, if the underlying distribution is  $S_B$  or Beta and decisions are made on the basis of an  $S_U$ , or if the true distribution is  $S_U$  and decisions are made using an  $S_B$  or a Beta model, the degree of error and its economic significance are likely to be substantial.

## Conclusion

A first general observation is that although the yield data used in the analyses is from the same state and crop, the SK combinations implied by the best fitting models are scattered over a large region of the SK plane corresponding to both the  $S_U$  and the  $S_B$ /Beta distributions, which suggests a need for candidate model alternatives that comprehensively span the SK space.

In addition, it is concluded that substantial, economically relevant errors in model fit should be expected if such alternatives are not considered and the assumed distribution is inconsistent with the SK profile of the true distribution underlying the data. Alternatively, if the assumed distribution is capable of accommodating the underlying MVSK values, errors due to discrepancies in higher-order moments or in the support characteristics of the assumed versus the true distribution appear to be relatively minor.

Following the recommended strategy of always considering the  $S_U$  and  $S_B$  distributions as potential candidate models could substantially

reduce the specification error risk that has long been associated with parametric methods, perhaps to an acceptable level in most applications. This conclusion, however, should be strengthened by further testing the performance of the  $S_U$ - $S_B$  family versus other parametric distributions, and does not preclude their consideration as candidate models. A particularly promising alternative for future testing is the multivariate normal mixture, which could also be reparameterized to span all MVSK combinations. It is also recognized that the relative complexity of the proposed family versus the most commonly used alternatives could affect its widespread applicability.

A final caveat on the recommended distributions is that they are continuous in nature. Thus, econometric modeling allowances must be made when working with data discontinuities such as censored yield observations due to droughts or flooding. It is also recognized that a statistically reliable use of the procedures discussed in this article requires at least moderate sample sizes (30–50 observations), which are often not available at the individual farm level. However, multivariate extensions of these procedures that can pool information from several farms to estimate the skewness and kurtosis parameters are feasible and straightforward, and county, state, and country level data now span several decades. Finally, as sample sizes grow, these procedures will become more usable at the single farm level as well.

## References

- Anderson, J.R. "Simulation: Methodology and Application in Agricultural Economics." *Review of Marketing and Agricultural Economics* 42(1974):3–55.
- Coble, K.H., T.O. Knight, R.D. Pope, and J.R. Williams. "Modeling Farm-level Crop Insurance Demand with Panel Data." *American Journal of Agricultural Economics* 78(1996): 439–47.
- Gallagher, P. "U.S. Soybean Yields: Estimation and Forecasting with Nonsymmetric Disturbances." *American Journal of Agricultural Economics* 69(1987):796–803.

- Johnson, N.L. "System of Frequency Curves Generated by Method of Translation." *Biometrika* 36(1949):149–76.
- Ker, A.P., and K. Coble. "Modeling Conditional Yield Densities." *American Journal of Agricultural Economics* 85(2003):291–304.
- Mood, A.M., F.A. Graybill, and D.C. Boes. *Introduction to the Theory of Statistics*. 3<sup>rd</sup> ed. New York: McGraw-Hill, 1974.
- Moss, C.B., and J.S. Shonkwiler. "Estimating Yield Distributions Using a Stochastic Trend Model and Non-Normal Errors." *American Journal of Agricultural Economics* 75(1993):1056–62.
- Nelson, C.H., and P.V. Preckel. "The Conditional Beta Distribution as a Stochastic Production Function." *American Journal of Agricultural Economics* 71(1989):370–78.
- Norwood, B., M.C. Roberts, and J.L. Lusk. "Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions." *American Journal of Agricultural Economics* 86(2004): 1032–43.
- Ramirez, O.A. "Estimation 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(1997):191–205.
- Ramirez, O.A. and T.U. McDonald. "Ranking Crop Yield Models: A Comment." *American Journal of Agricultural Economics* 88(2006): 1105–10.
- Ramírez, O.A., S.K. Misra, and J.E. Field. "Crop Yield Distributions Revisited." *American Journal of Agricultural Economics* 85(2003): 108–20.
- Ramirez, O.A., C.B. Moss, and W.G. Boggess. "Estimation and Use of the Inverse Hyperbolic Sine Transformation to Model Non-Normal Correlated Random Variables." *Journal of Applied Statistics* 21(1994):289–305.
- Taylor, C.R. "Two Practical Procedures for Estimating Multivariate Non-Normal Probability Density Functions." *American Journal of Agricultural Economics* 72(1990):210–17.
- Yatchew, A. "Nonparametric Regression Techniques in Economics." *Journal of Economic Literature* 36(1998):669–721.