Crop-Yield Distributions Revisited

Octavio A. Ramirez, Sukant Misra, and James Field

This article revisits the issue of crop-yield distributions using improved model specifications, estimation, and testing procedures that address the concerns raised in recent literature, which could have invalidated previous findings of yield nonnormality. It concludes that some aggregate and farm-level yield distributions are nonnormal, kurtotic, and right or left skewed, depending on the circumstances. The advantages of utilizing nonnormal versus normal probability distribution function models, and the consequences of incorrectly assuming crop-yield normality are explored.

Key words: crop-yields, Corn Belt, nonnormality, probability distributions, Texas cotton.

The issue of crop-yield distributions has been explored in the agricultural economics literature since the early 1970s, using two main types of procedures: parametric and nonparametric. Nonparametric methods are distribution free (Hogg and Craig), and have the advantage of being free from functional form and distributional assumptions. As such, they are impervious to specification error and might result in more accurate and robust models (Featherstone and Kastens). However, the nonparametric approach can be problematic when analyzing multiple variables with small samples. In agricultural economics, nonparametric procedures have been applied by Featherstone, Moghnieh, and Goodwin; Goodwin and Ker; and Ker and Goodwin among others.

Parametric methods require functional form and distributional assumptions and are, therefore, susceptible to specification errors and their statistical consequences. However, they work well under small sample conditions. Gallagher advances a univariate parametric procedure to model and simulate skewed yield distributions using the Gamma density.

Taylor approaches the problem of multivariate nonnormal modeling and simulation using a cubic polynomial approximation of a cumulative distribution function instead of assuming a particular density for empirical analysis. Ramirez, Moss, and Boggess explore the use of a multivariate nonnormal parametric modeling procedure, which is modified by Ramirez to analyze aggregate Corn Belt yields. He concludes that annual average Corn Belt corn and soybean yields (1950–89) are nonnormally distributed and left skewed. A consensus about the nonnormality of some crop-yield distributions, however, has not been reached in the agricultural economics literature, and recent research (Just and Weninger) points to model specification and statistical testing problems that shed doubt on the validity of all previous findings of yield nonnormality.

The following potential problems have been identified: (*a*) misspecification of the nonrandom components of the yield distributions, specifically, the assumption of linearity in the time trend for the mean of the distribution, and the ad hoc modeling of heteroskedasticity through arbitrarily specified dummy variables; (*b*) disregarding of the correlation between the yield variables involved in the normality tests; (*c*) conducting the normality tests under restricted model specifications; (*d*) misreporting of statistical significance, specifically using the results of separate (nonjoint) tests for skewness and kurtosis to conclude nonnormality; and (*e*) the use of aggregate time-series data to represent farm-level yield distributions and to estimate their variances.

There are also concerns about the inconsistency of the yield nonnormality findings, such as Day's reporting right skewness whereas others (Gallager, Swinton and King, Ramirez) conclude left skewness; and about the use of competing alternative distributional assumptions (Just and Weninger). The issue of whether a researcher conducting economic

---

Octavio Ramirez and Sukant Misra are associate professors and James Field is a former graduate student at the Department of Agricultural and Applied Economics, Texas Tech University.

The authors acknowledge the helpful comments and suggestions of Don Ethridge, Eduardo Segarra, Michael Livingston, two anonymous reviewers, and the feedback and guidance of Spiro Stefanou and an anonymous Associate Editor during the process of reviewing this research.

Amer. J. Agr. Econ. 85(1) (February 2003): 108–120

risk analyses should assume yield normality or allow for the possibility of yield nonnormality is critical. Distributional misspecification could seriously impact, for example, the results of crop insurance analyses, and nonnormality could invalidate mean-variance (E-V) approximations of expected utility maximization.

In this article we revisit the issue of yield nonnormality addressing all of the procedural problems discussed above. An expanded, refined parameterization of Johnson S_U family of densities (Johnson, Kotz, and Balakrishnan) is utilized, arguing that it is flexible enough to alleviate the concerns of using competing distributional assumptions in applied research. This expanded S_U family is used to revisit the issue of whether aggregate Corn Belt corn, soybean, and wheat yield distributions are nonnormal, relaxing the assumption of time-trend linearity, using joint tests for nonnormality under unrestricted model specifications to avoid the "double-jeopardy of normality" problem and ensure that the conclusions are not affected by the ordering of the statistical tests, or by ignorance of yield correlations. The tests are conducted under the most common heteroskedastic specification where the standard deviation is a function of time.

West Texas dryland cotton yields are also analyzed, illustrating the use of the expanded S_U family to jointly estimate county- and farm-level yield distributions. County- (TASS, 1970–99) and farm-level (TAES Farm Assist Program, 1988–97) data from five Texas Plains counties (Childress, Cochran, Crosby, Hale, and Wichita) and nine different farm units from two of these counties (five from Childress and four from Wichita county) are used to estimate the corresponding yield distributions. This article also provides explanations for the apparently contradictory findings of positively and negatively skewed crop-yield distributions.

Methods and Procedures

The S_U family of parametric distributions is built from a Gaussian density (Johnson, Kotz, and Balakrishnan). The S_U family can be modified and expanded by one parameter to obtain a flexible probability distribution function (pdf) model

$$(1)\quad Y_t = \mathbf{X}_t\mathbf{B} + \left[\{\sigma_t^2/G(\theta,\mu)\}^{1/2}\{\sinh(\theta V_t) - F(\theta,\mu)\}\right]/\theta,\ V_t \sim N(\mu,1),$$

$$F(\Theta,\mu) = E[\sinh(\theta V_t)] = \exp(\theta^2/2)\sinh(\theta\mu),$$

and

$$G(\Theta,\mu) = \{\exp(\theta^2)-1\}\{\exp(\theta^2) \times \cosh(-2\theta\mu)+1\}/2\theta^2$$

where Y_t is the random variable of interest (crop-yields); $\mathbf{X}_t$ is a $(1\times k)$ vector of exogenous variable values shifting the mean of the Y_t distribution through time (t); $\mathbf{B}$ is a $(k\times 1)$ vector of parameters; $\sigma_t^2>0, -\infty<\theta<\infty$, and $-\infty<\mu<\infty$, are other distributional parameters; and sinh, cosh, and exp denote the hyperbolic sine and cosine, and the exponential function, respectively. An independent normally distributed random variable, V_t, is the basis of the stochastic process defining the expanded S_U family of densities. From the results of Johnson, Kotz, and Balakrishnan (pp. 34–8) it follows that in this probability distribution function model

$$(2)\quad E[Y_t] = \mathbf{X}_t\mathbf{B},\qquad \mathrm{var}[Y_t]=\sigma_t^2,$$
$$\mathrm{skew}[Y_t] = S(\theta,\mu),\ \mathrm{kurt}[Y_t] = K(\theta,\mu)$$

where $S(\theta,\mu)$ and $K(\theta,\mu)$ involve combinations of exponential and hyperbolic sine and cosine functions. The former imply that $E[Y_t]=\mathbf{X}_t\mathbf{B}$, regardless of the values of σ_t^2, θ, and μ, and that the variance of Y_t is solely determined by σ_t^2. The skewness and kurtosis of the Y_t distribution are determined by the parameters θ and μ. If $\theta \neq 0$ and μ approaches zero, the Y_t distribution becomes symmetric, but it remains kurtotic. Higher absolute values of θ cause increased kurtosis. If $\theta \neq 0$ and $\mu > 0$, Y_t has a kurtotic and right-skewed distribution, whereas $\mu < 0$ results in a kurtotic and left-skewed distribution. Higher absolute values of μ produce increased skewness.

In practice, under normality, θ would approach zero and the proposed pdf model would collapse into a normal distribution with mean $\mathbf{X}_t\mathbf{B}$ and variance σ_t^2. Therefore, the null hypothesis of normality versus the alternative of asymmetric nonnormality is H_o: $\theta=\mu=0$ versus H_a: $\theta\neq 0$, $\mu\neq 0$. Note that because this is set up as a joint test for skewness and kurtosis, it does not suffer from the "double-jeopardy of normality" problem discussed in the recent literature (Just and Weninger). The null hypothesis of symmetric nonnormality versus the alternative of asymmetric nonnormality is H_o: $\theta\neq 0$, $\mu=0$ versus H_a: $\theta\neq 0$, $\mu\neq 0$.

Johnson, Kotz, and Balakrishnan (pp. 34–8) indicate that both the normal and the log-normal density are limiting cases of the S_U family, which also provides for a close approximation for the Pearson family of distributions. They demonstrate that the S_U, and, therefore, the expanded S_U family proposed in this study, allow for any combination of right or left skewness-leptokurtosis values below the log-normal line. This means that as long as the rare negative (platy) kurtosis can be ruled out, the expanded S_U family is flexible enough to alleviate the concerns of imposing incorrect distributional assumptions when using it to approximate a true, unknown crop-yield distribution. The concentrated log-likelihood function for the nonnormal pdf model defined in equation (1) is obtained using the transformation technique (Mood, Graybill, and Boes)

$$(3) \quad LL = \sum_{t=1}^{T} \ln(G_t) - 0.5 \times \sum_{t=1}^{T} H_t^2;$$

where

$$G_t = \{\sigma_t^2 / G(\theta, \mu)(1 + R_t^2)\}^{-1/2}$$

$$H_t = \{\sinh^{-1}(R_t)/\theta\} - \mu,$$

$$R_t = [\theta(Y_t - \mathbf{X}_t\mathbf{B}) / \{\sigma_t^2 / G(\theta, \mu)\}^{1/2}] + F(\theta, \mu)$$

where $t = 1, \ldots, T$ refers to the observations, $\sinh^{-1}(x) = \ln\{x + (1+x^2)^{1/2}\}$ is the inverse hyperbolic sine function, and σ_t^2, $F(\theta, \mu)$, and $G(\theta, \mu)$ are as defined in equation (1).

A multivariate equivalent is obtained by assuming that each of the M random variables of interest follows the flexible nonnormal pdf model defined in equation (1). All theoretically possible degrees of correlation among these variables are achieved by letting a multivariate normal vector $\mathbf{V}_t \sim N(\mu, \Sigma)$ underlie this model, where μ is an $(M \times 1)$ vector of parameters and Σ is an $(M \times M)$ correlation matrix with unit diagonal elements and nondiagonal elements ρ_{ij}. The log-likelihood function is obtained using the multivariate form of the transformation technique (Mood, Graybill, and Boes)

$$(4) \quad LL_M = \sum_{t=1}^{T} \sum_{j=1}^{M} \{\ln(G_{jt}) - 0.5[(\mathbf{H}_t \Sigma^{-1}) .{}^{*}\mathbf{H}_t]\} - 0.5T \ln(|\Sigma|)$$

where G_{jt} is as defined in equation (3) for each of the $j = 1, \ldots, M$ random variables of interest; $\mathbf{H}_t$ is a $1 \times M$ row vector with elements H_{jt} also as defined in equation (3).

The Empirical Models

The multivariate Corn Belt yield pdf model includes six parameters (θ_C, θ_S, θ_W, μ_C, μ_S, and μ_W) to account for potential corn (C), soybean (S), and wheat (W) nonnormality. Because of data limitations (thirty time-series observations per county times five counties, and ten observations per farm times nine farms), one set of nonnormality parameters is estimated for all county-level distributions and another for all farm-level distributions, that is, potential skewness and kurtosis in the Texas Plains county- and farm-level dryland cotton yield distributions are modeled by θ_{CL} and μ_{CL}, and θ_{FL} and μ_{FL}, respectively.

Both the full (nonnormal) and the restricted (normal) Corn Belt yield models are multivariate. They account for any contemporaneous yield correlations through the parameters ρ_{CS}, ρ_{CW}, and ρ_{SW}, eliminating a potential cause of inaccuracy in the statistical significance of the nonnormality tests discussed in recent literature. The Texas Plains cotton yield models are also multivariate. In the county-level model, ρ_{ij} accounts for any contemporaneous correlation between the yields in county i and county j. In the farm-level model, ρ_1 and ρ_2 account for the contemporaneous correlations between the yields in county 1 (Childress) and county 2 (Wichita) farms, respectively, whereas ρ_{12} accounts for possible correlation between farm-level yields across these two counties.

For the Corn Belt and Texas Plains county-level models, there are fifty and thirty annual observations available per crop/county. Thus, the means of the yield distributions are specified as fourth- and third-degree polynomial functions of time, respectively

$$(5) \quad \mathbf{X}_{jt}\mathbf{B}_j = B_{j0} + B_{j1}t + B_{j2}t^2 + B_{j3}t^3 + B_{j4}t^4,$$

and

$$\mathbf{X}_{jt}\mathbf{B}_j = B_{j0} + B_{j1}t + B_{j2}t^2 + B_{j3}t^3$$

where $j =$ C (corn), S (soybean), and W (wheat) in the Corn Belt model and $j = 1$ (Childress), 2 (Wichita), 3 (Crosby), 4 (Hale), 5 (Cochran) in the Texas Plains model; and t is a time-trend variable starting at $t = 1$. In both cases the standard deviations are initially

:cified as second-degree polynomial functions of time ($\sigma_{jt} = \sigma_{j0}+\sigma_{j1}t + \sigma_{j2}t^2$).

Since for the Texas Plains cotton farm-level models there are only ten years of yield data per farm unit, the means are specified as second-degree polynomial time trends only ($B_{j0}+B_{j1}t+B_{j2}t^2$, $j = 1,\dots,9$), and the variances of the yield distributions are assumed constant through time ($\sigma_{jt}^2 = \sigma_{jt}^2$, $j=1,\dots,9$).

The parameters determining the first four moments of and the correlations between the yield distributions are jointly estimated by maximum likelihood. This addresses another concern raised in recent literature that ignoring a critical distributional characteristic, that is, mean-trend nonlinearity, heteroskedasticity, or multivariate correlation, when testing for another, that is, nonnormality. invalidates the result of the test. This joint estimation and testing approach is preferable to the alternative used in previous studies of first modeling the mean, variance, and the correlation among distributions, and then using the detrended, heteroskedastic-corrected residuals to test for nonnormality, because the testing for time-trend nonlinearity and heteroskedasticity without accounting for potential nonnormality could affect the results of those tests. Also, full information procedures are more efficient than step-wise, limited-information procedures, which enhance the power of statistical tests on the models' parameters.

Results

Corn Belt Corn, Soybean, and Wheat Yield Distributions

The maximum likelihood parameter estimates for the multivariate nonnormal and normal Corn Belt yield pdf models are presented in table 1. The joint null hypothesis of normality ($H_o: \theta_C = \theta_S = \theta_W = \mu_C = \mu_S = \mu_W = 0$) is rejected in favor of the alternative hypothesis of nonnormality in at least one of the marginal pdfs at an exact $\alpha = 0.029$ level ($\chi^{2*}_{(6)} = -2[-279.22 - (-272.20)] = 14.04$). Note that this is a joint (skewness-kurtosis) multivariate (corn-soybean-wheat yield) test comparing a normal model that is nested to a nonnormal model, both of which include fourth-degree polynomial time trends on the means and second-degree polynomial time trends on the standard deviations of each of these three crop-yield distributions and account for the contemporaneous correlations between crop-yields. In other words, the null hypothesis of multivariate normality is rejected while addressing all of the concerns outlined in recent literature (Just and Weninger).

Analogous likelihood ratio tests (LRTs) for $H_o: \theta_C = \mu_C = 0$ ($\chi^{2*}_{(2)} = 7.53$) and $H_o: \theta_S = \mu_S = 0$ ($\chi^{2*}_{(2)} = 10.38$) (restricted models presented in table 1) separately reject normality in the Corn Belt corn and soybean yield distributions at $\alpha < 0.025$ and $\alpha < 0.010$ levels, respectively. The parameter estimates for θ_W and μ_W in the nonnormal model are equal to zero, indicating normality in the marginal distribution of wheat yields. Single-parameter LRTs (table 1) suggest that μ_C and μ_S are individually different from zero at the $\alpha < 0.05$ level, indicating that the corn and soybean yield distributions are skewed. The negative estimates for μ_C and μ_S imply left skewness, likely because of technological constraints imposing a ceiling to the maximum yields combined with the possibility of wide-spread drought or pest attack causing unusually low yields in any given year.

The final nonnormal and normal models are also presented in table 1. All of the parameters included in the final models are individually different from zero at an $\alpha < 0.15$ level of statistical significance, according to single-parameter LRTs, and the set of parameter restrictions leading from the full to the final models are not rejected at an $\alpha < 0.25$ level ($\chi^{2*}_{(11)} = 9.50$ in the nonnormal, and $\chi^{2*}_{(11)} = 7.96$ in the normal models). Thus, the final nonnormal and normal models are used in the following analyses.

The final models include second-, first-, and third-degree polynomial trends for the means and linear trends for the standard deviations of the corn, soybean, and wheat yield distributions, respectively, and a high-positive correlation between corn and soybean yields. The estimates for the parameters controlling the means and variances of the distributions, and their correlations, are similar under the normal and nonnormal models. As illustrated below, the main difference results from the nonnormality parameters.

The 1950–99 Corn Belt corn and soybean yield distributions are simulated numerically ($s = 500{,}000$ draws) using the parameter estimates from the final normal and nonnormal pdf models, a standard normal generator, and, in the case of the nonnormal model, the transformation to nonnormality specified in equation (1). Figures 1 and 2 illustrate the substantial left skewness of both distributions under the nonnormal pdf model. In comparison to the normal, both nonnormal yield

Table 1. Parameter Estimates for the Full and Restricted pdf Model Specifications for Corn Belt Corn, Soybean, and Wheat Yields

	Full Nonnormal Model	Full Normal Model	Rest Model 1	Rest Model 2	Final Nonnormal Model	Final Normal Model
MLLV	−272.20	−279.22	−275.96	−277.39	−276.95	−283.20
θ_C	0.4682[a]	0.0000	0.0000	0.3442	0.3681	0.0000
μ_C	−12.413[a]	0.0000	0.0000	−17.9483	−17.9410	0.0000
B_{C0}	49.4938[a]	47.3578	47.4567	48.4872	42.3748	43.1334
B_{C1}	−0.2658	0.6345	0.6325	0.1242	2.6611	2.5621
B_{C2}	24.3354[a]	17.6344	17.5337	21.0604	−1.7099	−1.5079
B_{C3}	−7.8128	−6.3886	−6.3099	−6.9274	0.0000	0.0000
B_{C4}	0.7453	0.6729	0.6591	0.6736	0.0000	0.0000
σ_{C0}	3.9091[b]	1.5202	1.3728	3.2610	3.4180	2.3872
σ_{C1}	0.0864	0.4026	0.4105	0.1517	0.2717	0.3103
σ_{C2}	0.5529	−0.2251	−0.2260	0.3343	0.0000	0.0000
θ_S	0.7341[a]	0.0000	0.5051	0.0000	0.5663	0.0000
μ_S	−0.9217[a]	0.0000	−1.4803	0.0000	−1.4423	0.0000
B_{S0}	22.4102[a]	21.6757	21.7802	22.2452	21.0292	20.9897
B_{S1}	−0.0823	0.1592	0.1525	−0.0451	0.4198	0.4213
B_{S2}	4.1268	2.4569	2.2538	4.0965	0.0000	0.0000
B_{S3}	−1.1893	−0.8228	−0.6794	−1.2711	0.0000	0.0000
B_{S4}	0.1117	0.0889	0.0672	0.1280	0.0000	0.0000
σ_{S0}	1.2595[a]	1.4321	1.3532	1.3199	1.4148	1.4108
σ_{S1}	0.0753[b]	0.0440	0.0387	0.0541	0.0431	0.0420
σ_{S2}	−0.0646	−0.0118	0.0096	−0.0281	0.0000	0.0000
B_{W0}	19.2701[a]	18.6720	18.7397	18.9870	19.3601	19.3601
B_{W1}	1.5660[a]	1.7589	1.7450	1.6546	1.4246	1.4246
B_{W2}	−5.4632[b]	−6.9693	−6.8876	−6.1677	−3.4947	−3.4947
B_{W3}	1.2078	1.6165	1.6011	1.4049	0.4236	0.4236
B_{W4}	−0.0924	−0.1279	−0.1272	−0.1101	0.0000	0.0000
σ_{W0}	3.5251[a]	3.3987	3.5199	3.4062	2.5855	2.5855
σ_{W1}	−0.0616	−0.0498	−0.0609	−0.0507	0.0670	0.0670
σ_{W2}	0.2707	0.2501	0.2687	0.2521	0.0000	0.0000
ρ_{CS}	0.6713[a]	0.7099	0.6783	0.6769	0.6843	0.7156
ρ_{CW}	0.2242	0.1731	0.1743	0.2128	0.0000	0.0000
ρ_{SW}	0.1497	0.1080	0.1473	0.1051	0.0000	0.0000

Notes: MLLV indicates the maximum value reached by the concentrated log-likelihood function. The parameter estimates corresponding to t^2, t^3, and t^4 are multiplied by 100, 1000, and 10000, respectively.

[a] Indicates that the parameter is statistically different from zero at the 5% level of statistical significance according to a likelihood ratio test.

[b] Indicates that the parameter is statistically different from zero at the 10% level of statistical significance according to a likelihood ratio test.

distributions allow for higher probabilities of relatively low yields and lower probabilities of relatively high yields, and their mass is heavily concentrated toward the upper side. The 1950–99 corn yield data are plotted in figure 3 versus the estimated second-degree polynomial trend and the 80% and 98% confidence bands for the yield realizations implied by the fifty corresponding yield distributions simulated under the normal model. These are obtained by identifying and joining the 10th and 90th and the 1st and 99th percentiles of the simulated distributions.

Note that, although none of the observed yields near the upper bound of the 98% confidence band, two yield realizations fall far below its lower bound. The confidence bands generated from the nonnormal model (figure 4) reflect the marked left skewness of the corn yield distributions depicted in figure 1. The 98% band contains all fifty yield realizations, as expected for this sample size. The number of observations found below the lower bound (Below) and above the upper bound (Above) of the 76% to 98% confidence bands versus the theoretically required numbers (TR) under the normal and nonnormal models are presented in table 2.

The normal model produces bounds that are clearly incompatible with the observed corn and soybean yield data. In the case of corn yields, for example, it leaves no observations above the upper boundary of its 88% band, while allowing for three yield realizations to

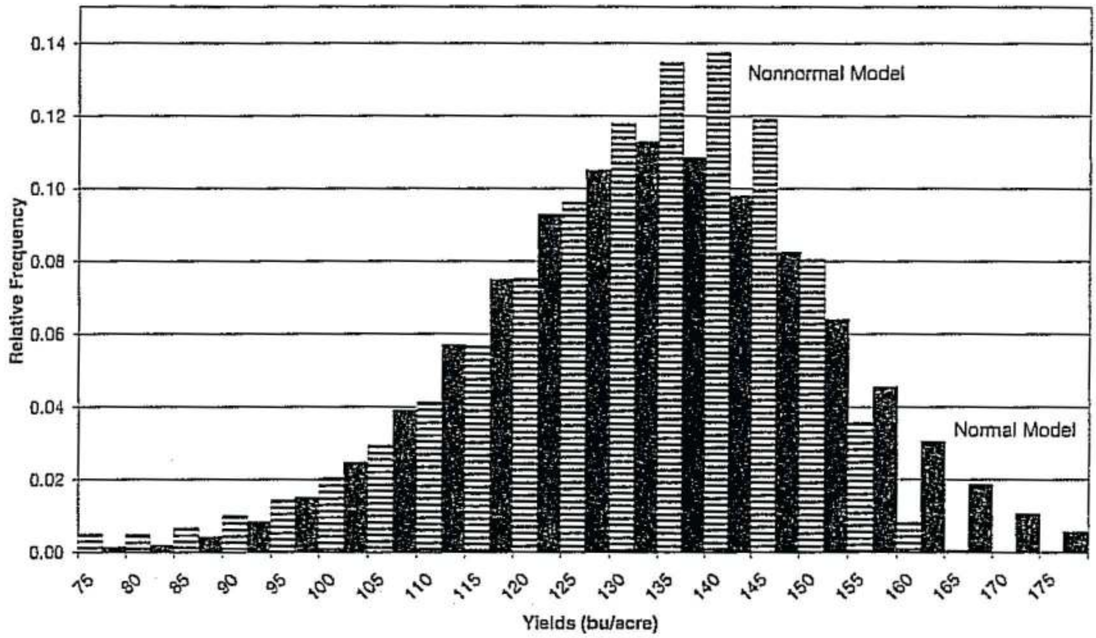

Figure 1. Simulated year 2000 Corn Belt corn yields under the normal and nonnormal models

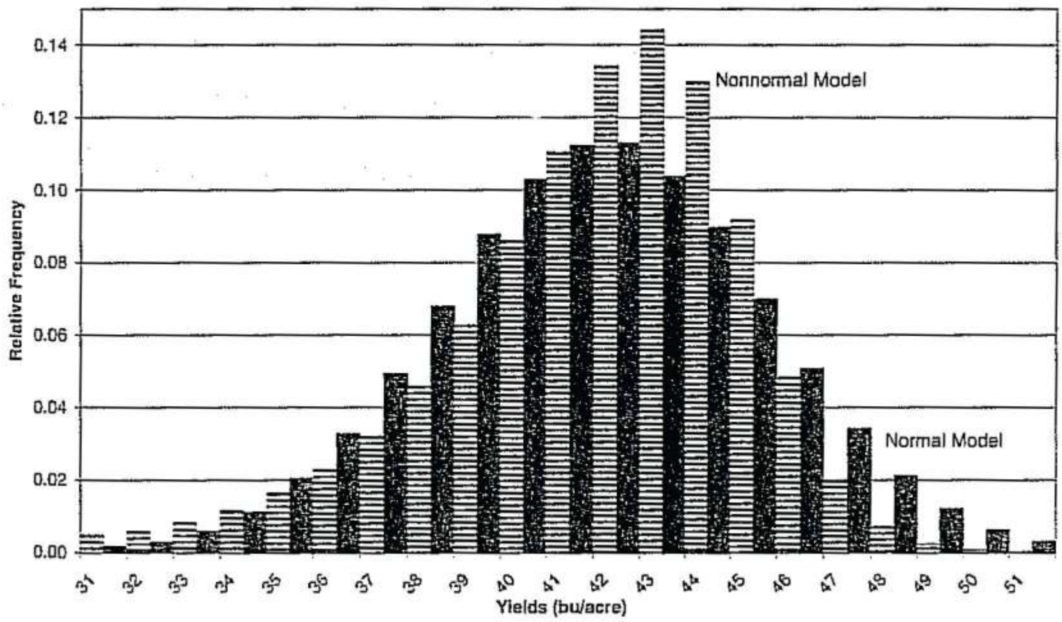

Figure 2. Simulated year 2000 Corn Belt soybean yields under the normal and nonnormal models

fall below the lower bound of its 92%, 94%, and 96% confidence bands and two under the 98% boundary. In the case of soybean yields, the normal model only leaves one observation above the upper boundary of its 78% band, while allowing for five yield realizations to fall below the lower bound of its 86% through 92% confidence bands, four under the 94%, and two under the 98% boundary. This overestimation of the lower bounds of the yield distributions by the normal model would be of particular concern for whether the simulated distributions were to be used for economic risk analysis.

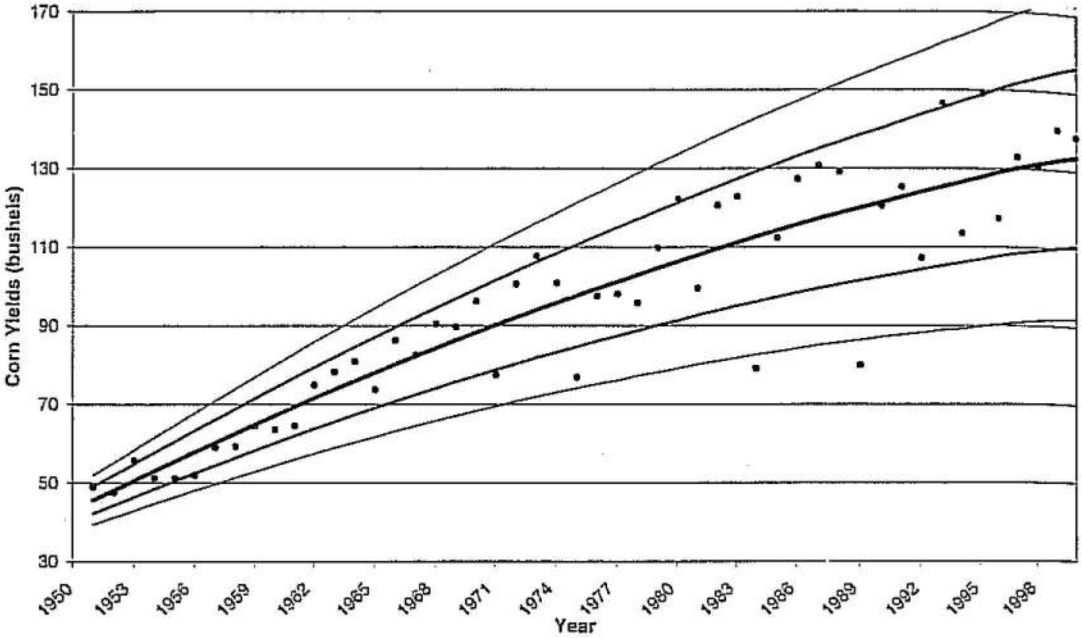

Figure 3. 80% and 98% confidence intervals for the corn yield realizations under the normal pdf model

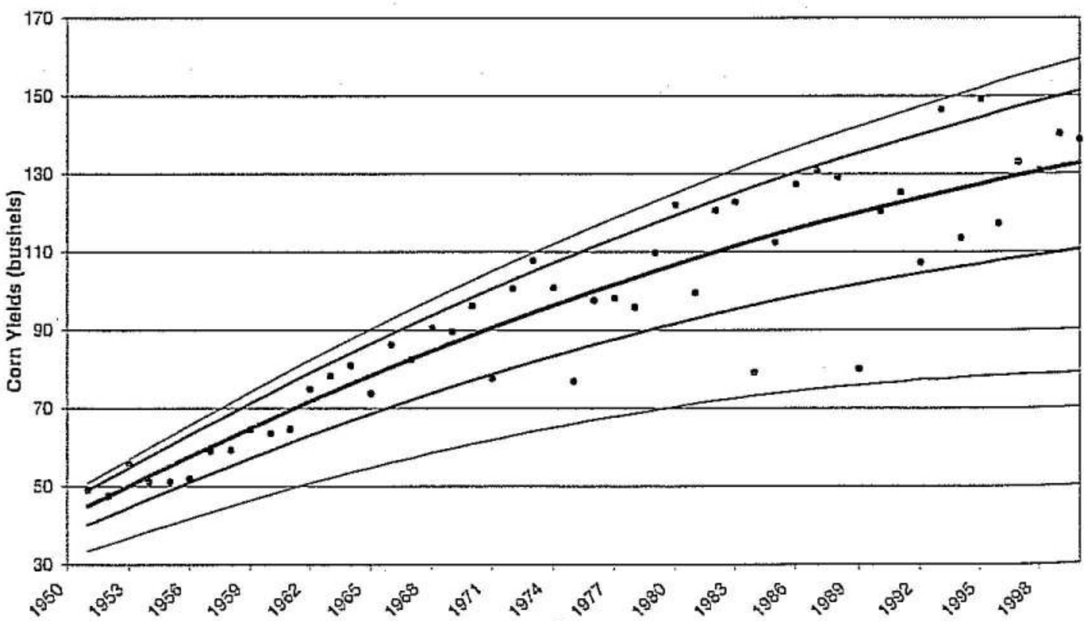

Figure 4. 80% and 98% confidence intervals for the corn yield realizations under the nonnormal pdf model

Although not perfect, the nonnormal model adheres better to the theoretically required numbers (table 2). It manages to better accommodate the data, particularly at confidence bands above 85%, by asymmetrically expanding the boundaries of the confidence bands as illustrated in figure 4. The normal model (figure 3), in contrast, implies unrealistically high upper bounds, whereas its lower bounds are still not low enough to account for the

Table 2. Select Statistics about the 76% to 98% Confidence Bands for the Corn Belt Corn and Soybean Yield Realizations under the Final Normal and Nonnormal pdf Models

%CI	TR	Below	Above	AW	Below	Above	AW
		Normal Corn Model			Nonnormal Corn Model		
76%	6.0	5	6	24.21	4	6	22.65
78%	5.5	5	5	25.27	4	6	23.71
80%	5.0	5	5	26.40	4	5	24.84
82%	4.5	5	3	27.62	4	5	26.09
84%	4.0	5	2	28.95	3	5	27.44
86%	3.5	3	1	30.40	3	5	28.96
88%	3.0	3	0	32.03	3	5	30.67
90%	2.5	3	0	33.88	3	5	32.65
92%	2.0	3	0	36.07	3	3	35.03
94%	1.5	3	0	38.76	2	1	38.02
96%	1.0	3	0	42.31	2	0	42.14
98%	0.5	2	0	47.93	0	0	49.01
		Normal Soybean Model			Nonnormal Soybean Model		
78%	5.5	6	1	6.09	7	6	5.44
80%	5.0	5	1	6.36	6	5	5.73
82%	4.5	5	1	6.65	5	5	6.04
84%	4.0	5	1	6.97	5	3	6.39
86%	3.5	5	1	7.32	5	3	6.78
88%	3.0	5	1	7.71	5	1	7.23
90%	2.5	5	1	8.16	4	1	7.76
92%	2.0	5	1	8.69	3	1	8.41
94%	1.5	4	0	9.33	2	1	9.26
96%	1.0	2	0	10.19	1	1	10.45
98%	0.5	2	0	11.55	0	1	12.55

Notes: TR refers to the number of observations that would be theoretically required to be below and above the boundaries of the confidence band; Below and Above are the actual numbers found below and above the lower and upper bounds, respectively; and AW stands for the average of the widths of the fifty confidence intervals comprising each of the bands.

lowest yield realizations. Also, as a result of the former, the confidence bands from the nonnormal model are narrower up to 95%, and on the average (table 2). The statistical significance of the nonnormal model is reflected on an improved representation of the observed 1950–99 Corn Belt corn and soybean yield data.

Texas Plains Dryland Cotton Yield Distributions

The parameter estimates for the Texas Plains county-level dryland cotton yield pdf models are presented in table 3. An LRT statistic comparing the full nonnormal with the normal model ($\chi^{2*}_{(2)} = -2[-688.92 - (-680.84)] = 16.16$) rejects the null hypothesis of normality (H_o: $\theta_{CL} = 0$, $\mu_{CL} = 0$) in favor of the alternative hypothesis of nonnormality (H_a: $\theta_{CL} \neq 0$, $\mu_{CL} \neq 0$) at a high level of statistical significance ($\alpha < 0.01$). A single-parameter LRT indicates that μ_{CL} is different from zero ($\alpha = 0.01$), suggesting that the county-level yield distributions are skewed. The positive μ_{CL} parameter estimate implies right skewness.

All of the parameters in the final models (table 3) are statistically different from zero at an $\alpha < 0.15$ level, according to single-parameter LRTs, and the set of parameter restrictions leading from the full to the final models are not rejected at an $\alpha < 0.25$ level ($\chi^{2*}_{(23)} = 27.98$ in the nonnormal and $\chi^{2*}_{(23)} = 26.47$ in the normal models). Thus, the final models are used in the analyses. They imply that county-level dryland cotton yields have remained constant in the Texas Plains during the last thirty years, a hypothesis maintained by many farmers, cotton experts, and industry groups. Surprisingly, yield variability appears to have increased in two of the five counties analyzed.

Cotton production in the Texas Plains is located mostly on the Southern Plains of West Texas, on an approximately 200 × 200 mile square comprising about thirty counties. The five counties in the analyses are spread across this area. The contemporaneous yield correlations between these counties range from 0.4 to 0.85 and are statistically significant at an $\alpha < 0.01$ level, with the exception of

Table 3. Parameter Estimates for Full and Restricted pdf Model Specifications for Dryland Cotton Yields in Five Texas Southern Plains Counties

	Full Nonnormal				Full Normal				Final Nonnormal		Final Normal
MLLV	−680.84				−688.92				−694.82		−702.16
θ_{CL}	0.4298[a]	σ_{01}	105.987[a]	θ_{CL}	0.0000	σ_{01}	79.8599	θ_{CL}	0.2912	θ_{CL}	0.0000
μ_{CL}	19.0872[a]	σ_{02}	34.7798[a]	μ_{CL}	0.0000	σ_{02}	53.0573	μ_{CL}	20.5540	μ_{CL}	0.0000
B_{01}	312.153[a]	σ_{03}	89.3634[a]	B_{01}	311.519	σ_{03}	131.131	B_{01}	304.8850	B_{01}	307.0186
B_{02}	178.871[a]	σ_{04}	52.3243[a]	B_{02}	179.431	σ_{04}	67.7838	B_{02}	168.5700	B_{02}	172.6908
B_{03}	361.003[a]	σ_{05}	0.0000	B_{03}	373.126	σ_{05}	0.0000	B_{03}	259.9345	B_{03}	261.5790
B_{04}	219.606[a]	σ_{11}	12.8549	B_{04}	255.357	σ_{11}	14.5045	B_{04}	222.9672	B_{04}	227.1613
B_{05}	197.943[a]	σ_{12}	17.3445	B_{05}	190.212	σ_{12}	14.8010	B_{05}	228.6126	B_{05}	230.0757
B_{11}	−12.6773	σ_{13}	5.3478	B_{11}	−10.4489	σ_{13}	−3.7842	σ_{01}	128.1141	σ_{01}	125.1818
B_{12}	−6.3333	σ_{14}	3.5381	B_{12}	3.3094	σ_{14}	5.5027	σ_{02}	123.1535	σ_{02}	130.4016
B_{13}	−30.8387	σ_{15}	19.7696[a]	B_{13}	−25.3250	σ_{15}	18.3018	σ_{03}	100.7046	σ_{03}	95.7795
B_{14}	1.6568	σ_{21}	−5.3610	B_{14}	−10.0087	σ_{21}	−5.5928	σ_{04}	69.7763	σ_{04}	83.1788
B_{15}	−0.4673	σ_{22}	−5.2674	B_{15}	6.9081	σ_{22}	−4.8539	σ_{05}	75.6322	σ_{05}	76.5918
B_{21}	1.8333	σ_{23}	−2.0949	B_{21}	1.5088	σ_{23}	0.4602	σ_{15}	2.0086	σ_{15}	1.8723
B_{22}	1.0542	σ_{24}	0.8445[a]	B_{22}	−0.0325	σ_{24}	−1.3130	σ_{24}	1.4111	σ_{24}	0.9460
B_{23}	2.4133	σ_{25}	−6.2376	B_{23}	1.6277	σ_{25}	−5.6941	ρ_{21}	0.5176	ρ_{21}	0.5533
B_{24}	−0.1970	ρ_{21}	0.5450[a]	B_{24}	0.8283	ρ_{21}	0.6138	ρ_{31}	0.6173	ρ_{31}	0.6121
B_{25}	0.9501	ρ_{31}	0.5994[a]	B_{25}	0.1143	ρ_{31}	0.6557	ρ_{32}	0.6980	ρ_{32}	0.7059
B_{31}	−5.3610	ρ_{32}	0.7760[a]	B_{31}	−4.4518	ρ_{32}	0.7656	ρ_{41}	0.4009	ρ_{41}	0.4356
B_{32}	−3.3254	ρ_{41}	0.4512[a]	B_{32}	−0.6756	ρ_{41}	0.5681	ρ_{42}	0.6463	ρ_{42}	0.6540
B_{33}	−5.2814	ρ_{42}	0.7438[a]	B_{33}	−3.2014	ρ_{42}	0.7143	ρ_{43}	0.5982	ρ_{43}	0.5990
B_{34}	0.8146	ρ_{43}	0.6608[a]	B_{34}	−1.7598	ρ_{43}	0.6627	ρ_{51}	0.8461	ρ_{51}	0.8540
B_{35}	−3.6913	ρ_{51}	0.8628[a]	B_{35}	−1.5490	ρ_{51}	0.8800	ρ_{52}	0.6203	ρ_{52}	0.6400
		ρ_{52}	0.6377[a]			ρ_{52}	0.6833	ρ_{53}	0.4909	ρ_{53}	0.4960
		ρ_{53}	0.5005[a]			ρ_{53}	0.5579	ρ_{54}	0.2202	ρ_{54}	0.2731
		ρ_{54}	0.3335[b]			ρ_{54}	0.4556				

Notes: MLLV indicates the maximum value reached by the concentrated log-likelihood function. All parameters in the final models are statistically significant at the 10% level, with the exception of ρ_{54}, which is significant at the 15% level. The parameter estimates corresponding to r^2, r^3, and r^4 are multiplied by 100, 1000, and 10000, respectively.

[a] Indicates that the parameter is statistically different from zero at the 5% level of statistical significance according to a likelihood ratio test.

[b] Indicates that the parameter is statistically different from zero at the 10% level of statistical significance according to a likelihood ratio test.

Hale-Cochran's which is 0.22 and only significant at an $\alpha = 0.15$ level. An LRT indicates that not all correlation coefficients are statistically equal. However, in contrast to what has been hypothesized in previous studies (Just and Weninger), within the area under analysis, increased distance between counties does not appear to decrease the degree of correlation between their yields.

As in the case of Corn Belt yields, the mean and standard deviation parameters estimated under the final normal and nonnormal models are quite similar. The residuals and the standard deviation parameters from the normal model are used to obtain $n = 5 \times 29 = 145$ standardized residuals that would be theoretically drawn from a distribution with mean zero and variance one. These residuals are multiplied by the estimated standard deviation and added to the estimated mean for Wichita county to obtain 145 adjusted yield observations. The relative frequency distribution of this adjusted yield data is compared to the yield distributions for Wichita county implied by the normal and nonnormal models, simulated using the same procedures described for Corn Belt yields.

The normal model clearly overestimates the probability of very low yields (below 110 lbs/acre) underestimates the probability of moderately low-to-average yields (110 to 230 lbs/acre), overestimates the probability of average-to-moderately-high yields (230 to 310 lbs/acre), and underestimates the probability of very high yields (above 390 lbs/acre) (figure 5). The nonnormal model is more accurate than the normal on predicting the observed yield frequencies in twelve of the thirteen intervals depicted in figure 5, and does not show substantial under- or overestimation patterns. The average of the absolute differences between the observed relative frequencies and those predicted using the simulated distributions is 0.076 under the nonnormal model,

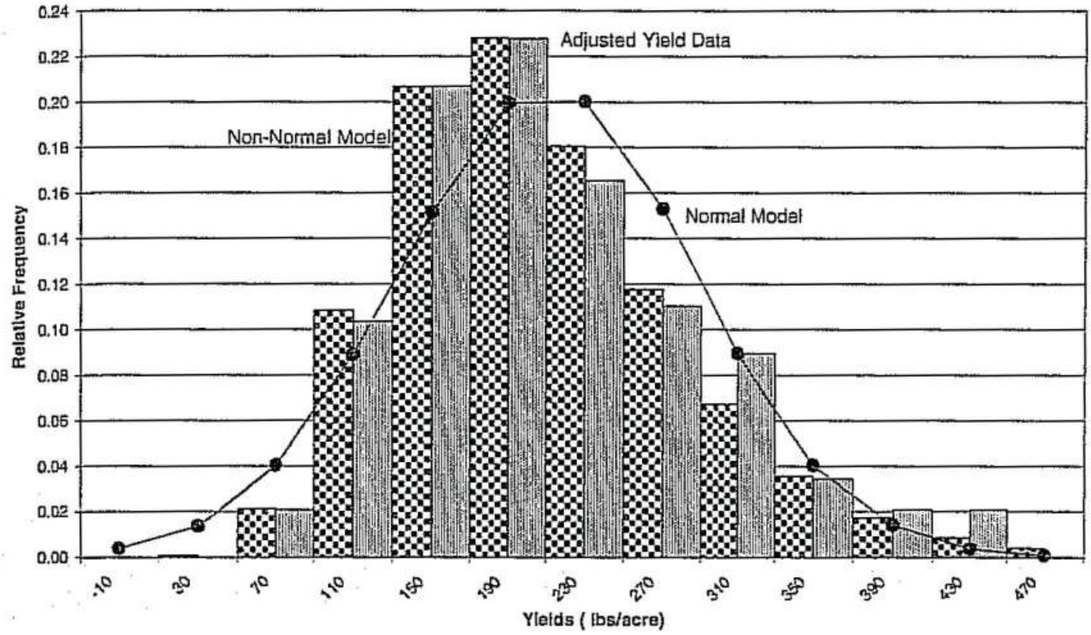

Figure 5. Simulated west Texas county-level dryland yields under the normal and nonnormal models versus adjusted data

versus 0.244 under the normal model. Because the nonnormal model assumes the same kewness and kurtosis parameters for all five counties, similar results would be expected if adjusting the standardized residuals by the estimated mean and standard deviation of any other county and comparing them with the yield distributions simulated for that particular county.

The maximum likelihood parameter estimates for the Texas Plains farm-level dryland cotton yield pdf models for Childress and Wichita counties are presented in table 4. An LRT statistic comparing the full nonnormal with the normal model $\chi^{2*}_{(2)} = -2[-466.54 - (-461.61)] = 9.85$) rejects the null hypothesis of normality (H_o: $\theta_{FL} = 0$, $\mu_{FL} = 0$) in favor of the alternative hypothesis of nonnormality (H_a: $\theta_{FL} \neq 0$, $\mu_{FL} \neq 0$) at a high level of statistical significance ($\alpha < 0.01$). A single-parameter LRT test indicates that μ_{FL} is individually different from zero at the $\alpha = 0.01$ level, suggesting that the farm-level cotton yield distributions are skewed. As in the case of the county-level distributions, the positive value of the μ_{FL} parameter estimate implies right-skewness.

The yield right skewness is compatible with Texas Southern Plains farmers' and researchers' intuition: Given normal rainfall conditions of 8–12 inches during the critical (May to August) period of the growing season, dryland cotton production systems have evolved to produce between 150 and 350 lbs/acre, (200–250 lbs/acre, on average, depending on the county). Under minimum rainfall (4–6 inches) that occurs about once a decade, many farms report low (50–150 lbs/acre) yields per harvested acre. Extremely favorable rainfall amounts of 15–20 inches occur in certain areas every 10–15 years, resulting in yields of between 400 and 600 lbs/acre.

In other words, the right skewness of the dryland cotton yield distribution is likely derived from the right skewness of the rainfall distribution. Including rainfall as a factor shifting the mean of the yield distribution from year to year could result in a conditional yield distribution that is normal. This, however, would be conditional on prior knowledge of the amount of rainfall that would occur in any given year, which is not compatible with the usual risk analyses applications of simulated yield distributions.

The final nonnormal and normal models (table 4) are formulated considering the results of the single-parameter LRTs on the remaining coefficients. All of the parameters included in these models are individually different from zero at an $\alpha < 0.15$ level of statistical significance, according to single-coefficient

Table 4. Parameter Estimates for Full and Restricted pdf Model Specifications for Dryland Cotton Yields from Nine Farms in Childress and Wichita Counties

Full Nonnormal				Full Normal				Final Nonnormal		Final Normal	
MLLV	−461.61				−466.54				−465.91		−471.03
θ_{FL}	0.5252[a]	σ_{01}	154.395[a]	θ_{FL}	0.0000	σ_{01}	156.171	θ_{FL}	0.4765	θ_{FL}	0.0000
μ_{FL}	2.5171[a]	σ_{02}	93.5051[a]	μ_{FL}	0.0000	σ_{02}	94.445	μ_{FL}	2.8194	μ_{FL}	0.0000
B_{01}	577.2013[a]	σ_{03}	147.8967[a]	B_{01}	635.125	σ_{03}	150.079	B_{01}	545.574	B_{01}	622.1803
B_{02}	281.4454[a]	σ_{04}	295.8783[a]	B_{02}	255.642	σ_{04}	275.191	B_{02}	250.912	B_{02}	257.8000
B_{03}	760.9779[a]	σ_{05}	191.613[a]	B_{03}	684.236	σ_{05}	164.901	B_{03}	728.003	B_{03}	671.8015
B_{04}	756.4156[a]	σ_{06}	80.2997[a]	B_{04}	666.572	σ_{06}	75.630	B_{04}	374.261	B_{04}	380.5999
B_{05}	347.309[a]	σ_{07}	100.5253[a]	B_{05}	281.321	σ_{07}	92.697	B_{05}	337.405	B_{05}	342.7999
B_{06}	334.8997[a]	σ_{08}	226.2606[a]	B_{06}	293.612	σ_{08}	171.140	B_{06}	293.235	B_{06}	277.4821
B_{07}	354.6907[a]	σ_{09}	93.1713[a]	B_{07}	352.201	σ_{09}	73.762	B_{07}	322.331	B_{07}	335.2046
B_{08}	486.2828[a]	ρ_{11}	0.3185[b]	B_{08}	335.497	ρ_{11}	0.231	B_{08}	304.717	B_{08}	295.8000
B_{09}	298.4769[a]	ρ_{22}	0.7163[a]	B_{09}	237.282	ρ_{22}	0.702	B_{09}	340.191	B_{09}	341.1000
B_{11}	−146.260[b]	ρ_{12}	0.1702	B_{11}	−147.57	ρ_{12}	0.133	B_{11}	−139.19	B_{11}	−155.861
B_{12}	−3.6809			B_{12}	11.2827			B_{13}	−136.42	B_{13}	−95.4052
B_{13}	−143.949[a]			B_{13}	−87.449			B_{16}	−13.441	B_{16}	−10.1967
B_{14}	−151.060			B_{14}	−82.322			B_{17}	−14.871	B_{17}	−16.5281
B_{15}	34.1359			B_{15}	69.2175			B_{21}	12.5497	B_{21}	13.2299
B_{16}	−37.8884			B_{16}	−12.604			B_{23}	8.8710	B_{23}	4.8007
B_{17}	−36.7196			B_{17}	−18.092			σ_{01}	151.973	σ_{01}	159.0956
B_{18}	−79.1746			B_{18}	8.2339			σ_{02}	100.880	σ_{02}	96.5063
B_{19}	14.5752			B_{19}	51.9613			σ_{03}	148.420	σ_{03}	152.6977
B_{21}	12.7250[b]			B_{21}	11.709			σ_{04}	306.671	σ_{04}	296.5955
B_{22}	−0.3429			B_{22}	−1.556			σ_{05}	215.981	σ_{05}	184.2491
B_{23}	9.0406[b]			B_{23}	3.341			σ_{06}	78.9314	σ_{06}	75.8993
B_{24}	11.5491			B_{24}	4.332			σ_{07}	96.4512	σ_{07}	93.0169
B_{25}	−5.2206			B_{25}	−8.291			σ_{08}	226.610	σ_{08}	176.3552
B_{26}	2.3337			B_{26}	−0.075			σ_{09}	97.9251	σ_{09}	82.1173
B_{27}	2.2049			B_{27}	−0.218			ρ_{11}	0.3750	ρ_{11}	0.2668
B_{28}	6.3834			B_{28}	−2.207			ρ_{22}	0.6643	ρ_{22}	0.6708
B_{29}	−1.0969			B_{29}	−4.726			ρ_{12}	0.1597	ρ_{12}	0.1602

Notes: MLLV indicates the maximum value reached by the concentrated log-likelihood function. All parameters in the final models are statistically significant at the 10% level, with the exception of ρ_{12}.

[a] Indicates that the parameter is statistically different from zero at the 5% level of statistical significance according to a likelihood ratio test.

[b] Indicates that the parameter is statistically different from zero at the 10% level of statistical significance according to a likelihood ratio test.

LRTs, and the set of parameter restrictions leading from the full to the final models are not rejected at an $\alpha < 0.25$ level ($\chi^{2*}_{(12)} = 8.61$ in the nonnormal, and $\chi^{2*}_{(12)} = 8.99$ in the normal models). Thus, the final nonnormal and normal models are used in the following analyses.

Unlike the county-level models, four of the nine final farm-level models show statistically significant time-trend parameters suggesting that the yields have declined during the 1988–97 period. This is likely not indicative of a real downward trend, but rather because of two very dry growing seasons occurring near the end of that period. As expected, the variances of the county-level distributions generally are lower than the variances of the farm-level distributions. The skewness and kurtosis parameters are quite different, as well. Statistically significant contemporaneous correlations of 0.38 and 0.66 are detected between the yields in Childress and in Wichita county farms, respectively.

Adjusted farm-level yields and the simulated yield distributions under the normal and nonnormal models are obtained for one of the farms in Childress county, following the same procedures used at the county level (figure 6). As in the case of county-level yields, the normal model overestimates the probability of very low yields (below 70 lbs/acre), underestimates the probability of moderately low-to-average yields (140 to 280 lbs/acre), overestimates the probability of average-to-moderately-high yields (280 to 420 lbs/acre), and underestimates the probability of very high yields (above 420 lbs/acre). The nonnormal model is more accurate than the normal model in predicting the observed

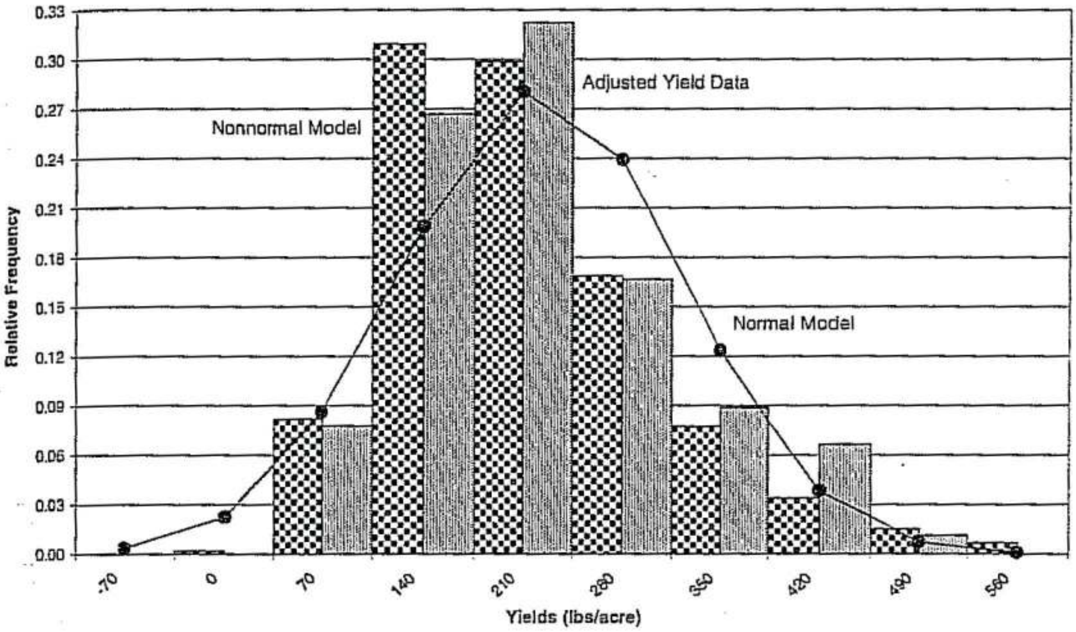

Figure 6. Simulated west Texas farm-level dryland yields under the normal and nonnormal models versus adjusted data

ive yield frequencies in eight of the ten i. ..rvals depicted in figure 6.

The average of the absolute differences between the observed relative frequencies and those predicted from the simulated distributions is 0.135 under the nonnormal model, versus 0.284 under the normal model. As in the county level, because the nonnormal model assumes the same skewness and kurtosis parameters for all nine farms, similar results would be expected if adjusting the standardized residuals by the estimated mean and standard deviation of any other farm and comparing them with the yield distributions simulated for that particular farm.

Conclusions and Recommendations

This article reaffirms Ramirez's (1997) findings that Corn Belt corn and soybean yields are nonnormally distributed and left skewed, using an expanded data set and addressing the procedural issues that have been raised in recent literature. The procedures used here are preferable to previous methods because they allow for the testing of all potential distributional characteristics (nonlinear trends in means, heteroskedasticity, kurtosis, right .eft skewness, and cross-distribution correlation) in a joint, full-information context, which is the most efficient. The tests for nonlinear trends and heteroskedasticy are conducted while allowing for any potential nonnormality, and vice versa, using the additional information transmitted through the cross-distribution correlation matrix.

As recognized by the authors of previous studies, nonrejection does not prove yield normality, because the magnitudes of the type-two errors in their normality tests are unknown. In contrast, here Corn Belt corn and soybean yields are found to be nonnormally distributed, with a small 3.0% probability of making an error in this conclusion. The consistency of the results after adding a substantial amount of recent data, and under an alternative, more common heteroskedastic specification, is further evidence of the soundness of the nonnormality concussions.

In the case of the Texas Plains dryland cotton yields, the normality hypothesis is rejected at the 1% significance level at both the farm and county levels, providing further support for the thesis that some crop-yield distributions are nonnormal. There is no contradiction in the findings of Corn Belt corn and soybean yield distribution left skewness and Texas Plains dryland cotton yield distribution right skewness. As argued above in detail, diverse nonnormality patterns could result from different critical factors affecting aggregate and farm-level yields, depending on the crop, cropping system, and geographical region.

The main recommendation of this study is that researchers estimating and simulating farm. county, state, regional, or U.S.-level yield distributions for policy, market, industry, farm, or any other type of risk analysis, should recognize that they could be nonnormal, and use the methods available for testing, and for estimating and simulating nonnormal distributions when necessary.

[Received February 2002; final revision received May 2002.]

References

Day, R.H. "Probability Distributions of Field Crop Yields." *Journal of Farm Economics* 47(1965): 713–41.

Featherstone, A.M. and T.L. Kastens. "Non-Parametric and Semi-Parametric Techniques for Modeling and Simulating Correlated, Non-Normal Price and Yield Distributions: Applications to Risk Analysis in Kansas Agriculture." *Journal of Agricultural and Applied Economics* 32(2)(2000):267–81.

Featherstone, A.M., G.A. Moghnieh and B.K. Goodwin. "Farm-Level Non-Parametric Analysis of Cost-Minimization and Profit-Maximization Behavior." *Agricultural Economics* 13(1995):111–20.

Gallagher, P. "U.S. Soybean Yields: Estimation and Forecasting with Nonsymmetric Disturbances." *American Journal of Agricultural Economics* 71(1987):796–803.

Goodwin, B.K. and A.P. Ker. "Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group Risk (GRP) Crop Insurance Contracts." *American Journal of Agricultural Economics* 80(1998):139–53.

Hogg, R.V. and A.T. Craig. *Introduction to Mathematical Statistics*. New York: McMillan, 1963.

Johnson, N.L., S. Kotz and N. Balakrishnan. *Continuous Univariate Distributions*. New York: Wiley & Sons, 1994.

Just, R.E. and Q. Weninger. "Are Crop Yields Normally Distributed?" *American Journal of Agricultural Economics* 81(1999):287–304.

Ker, A.P. and B.K. Goodwin. "Nonparametric Estimation of Crop Insurance Rates Revisited." *American Journal of Agricultural Economics* 83(2000):463–78.

Mood, A.M., F.A. Graybill and D.C. Boes. *Introduction to the Theory of Statistics*. New York: McGraw-Hill, 1974, pp. 198–212.

Ramirez, O.A. "Estimation and Use of a Multivariate Parametric Model for Simulating Heteroscedsatic, Correlated, Non-Normal Random Variables: The Case of Corn Belt Corn, Soybean and Wheat Yields." *American Journal of Agricultural Economics* 79(1997):291–305.

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):289–303.

Swinton, S. and R.P. King. "Evaluating Robust Regression Techniques for Detrending Crop Yield Data with Non-Normal Errors." *American Journal of Agricultural Economics* 73(1991):446–61.

Texas Agricultural Extension Service (TAES). Data obtained from Craig Fincham, TAES Economist, District 2. 11 November, 1999.

Taylor, C.R. "Two Practical Procedures for Estimating Multivariate Nonnormal Probability Density Functions." *American Journal of Agricultural Economics* 72(1990):210–17.

Texas Agricultural Statistics Service (TASS). Texas Agricultural Statistics reports, 1970–99.