**Are Crop Yields Normally Distributed?**

by

Octavio A. Ramirez[1,2,3,4]

Sukant Misra[1]

and

James Field[1]

[1] Associate Professor and MS student, respectively. Department of Agricultural and Applied Economics, Texas Tech University, Box 42132, Lubbock, TX 79409-2132. Email (senior author): Octavio.Ramirez@ttu.edu.

[2] The authors acknowledge the helpful comments and suggestions of Don Ethridge, Eduardo Segarra and Michael Livingston.

[3] Copyright 2001 by Octavio A. Ramirez. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

[4] Paper presented at the annual meeting of the American Agricultural Economics Association, Chicago, Illinois, August 5-8, 2001

# Are Crop Yields Normally Distributed?


## Abstract

This paper revisits the issue of crop yield distributions using improved model specifications, estimation and testing procedures that address the methodological concerns raised in recent literature that could have invalidated previous conclusions of yield non-normality. It shows beyond reasonable doubt that some crop yield distributions are non-normal, kurtotic and right or left skewed, depending on the circumstances. A procedure to jointly estimate non-normal farm- and aggregate-level yield distributions with similar means but different variances is illustrated, and the consequences of incorrectly assuming yield normality are explored.


**Key Words:** Yield non-normality, probability distribution function models, Corn Belt yields, West Texas dryland cotton yields.

**Are Crop Yields Normally Distributed?**

The issue of crop yield normality versus non-normality has been sporadically addressed in the agricultural economics literature since the early 1970s. In 1974 Anderson argues about the importance of being able to model crop yield non-normality (skewness and kurtosis) and changing variances in the yield distributions through time/space; since these could be important characteristics of many crop yield distributions and could have substantial implications for economic risk analyses.

Gallagher (1987) advances a univariate procedure to model and simulate skewed yield distributions using the Gamma density, focusing on modeling the changing variability of soybean yields over time as. He recognizes, however, that there are fixed relationships between the mean, the variance and the level of skewness and kurtosis imposed by the Gamma density, which depends on two parameters only. A consequence of this "lack of flexibility," for example, is that in order to model and simulate a changing variance one needs to accept that the mean, skewness and kurtosis of the yield distribution are also changing according to arbitrarily fixed formulae.

In 1990, Taylor tackles for the first time the problem of multivariate non-normal simulation. He uses a cubic polynomial approximation of a cumulative distribution function instead of assuming a particular multivariate density for empirical analysis. Ramirez, Moss and Boggess explore the use of a parametric density based on an inverse hyperbolic sine transformation to normality. Ramirez (1997) analyzes aggregate Corn Belt yields using a multivariate non-normal parametric modeling procedure. He concludes that annual average Corn Belt corn and soybean yields (1950-1989) are non-normally distributed with a tendency towards left-skewness.

A consensus about the possible non-normality 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 non-normality. The following specific problems have been identified: (*i*) misspecification of the non-random components of the yield distributions, more specifically the assumption of linearity in the time trend for the mean of the distribution, (*ii*) misreporting of statistical significance, more specifically using the results of separate (non-joint) tests for skewness and kurtosis to conclude non-normality, and (*iii*) the use of aggregate time series data to represent farm-level yield distributions, more specifically to estimate the variance of the farm-level yield distribution. In addition, there are concerns about the inconsistency of the yield non-normality findings, such as Day's reporting positive (right) skewness while others (Gallager, 1986, 1987; Swinton and King; Ramirez, 1997) conclude negative (left) skewness; and about the using of competing alternative distributional assumptions.

The issue of whether an applied researcher conducting economic risk analyses should assume yield normality or allow for the possibility of yield non-normality is critical. Distributional misspecification could fundamentally impact, for example, the results of crop insurance analyses, and non-normality could invalidate mean-variance (E-V) approximations of expected utility maximization (Just and Weninger). This article revisits the issue of yield non-normality while addressing all of the procedural problems discussed above.

Specifically, an expanded, more refined parameterization of Johnson $S_U$ family of densities is utilized, arguing that this parameterization is flexible enough to alleviate the concerns of using different competing distributional assumptions in applied research. This expanded $S_U$ family of densities is used to revisit the issue of whether the aggregate Corn Belt corn and soybean yield distributions are non-normal, relaxing the assumption of linearity in the time trends for the means on the distributions, using joint tests for non-normality under the full and all restricted 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. The tests are conducted under two different heteroskedastic specifications to explore if this could affect the non-normality conclusions.

West Texas dryland cotton yield distributions are also analyzed in this study, illustrating the use of the expanded $S_U$ family of densities to jointly estimate aggregate and farm-level yield distributions. A combination of county- (1970-1998) and farm- (1988-1998) level data from six Southern High Plain counties and 15 different farm units, and four Northern Low Plain counties and 10 different farm units, is used to estimate the corresponding distributions. This addresses another important issue recently raised in the literature: how to estimate different farm and aggregate level yield variance structures without assuming normality. The article also provides likely explanations for the apparently contradictory findings of positively and negatively skewed crop yield distributions, the last issue recently cited as evidence against the proposition of non-normality.

**Methods and Procedures**

The $S_U$ family of parametric distributions was 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) $Y_t = X_tB + [\{\sigma_t/G(\theta,\mu)\}^{1/2}\{\sinh(\theta V_t) - F(\theta,\mu)\}]/\theta, \;\; V_t \sim N(\mu,1),$

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

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

where $Y_t$ is the random variable of interest (crop yields); $X_t$ is a $(1xk)$ vector of exogenous variable values shifting the mean of the $Y_t$ distribution through time (t); B is a $(kx1)$ vector of parameters; $\sigma_t > 0$, $-\infty<\theta<\infty$, and $-\infty<\mu<\infty$, are other distributional parameters; and sinh, cosh, and exp denote the hyperbolic sine and cosine and the exponential function, respectively. $V_t$**,** an independent normally distributed random variable, is the basis of the stochastic process defining the expanded $S_U$ family of densities. Using the results of Johnson, Kotz and Balakrishnan it can be shown that in this model:

(2) $E[Y_t] = X_tB,$

$Var[Y_t] = \sigma_t,$

$Skew[Y_t] = E[Y_t^3] = S(\theta,\mu),$

$Kurt[Y_t] = E[Y_t^4] = K(\theta,\mu).$

where $S(\theta,\mu)$ and $K(\theta,\mu)$ involve combinations of exponential and hyperbolic sine and cosine functions. The results in (2) imply that $E[Y_t] = X_tB$, regardless of the values of $\sigma_t$, $\theta$, and $\mu$, and that the variance of $Y_t$ is solely determined by $\sigma_t$. The skewness and kurtosis of the $Y_t$ distribution are determined by the parameters $\theta$ and $\mu$. If $\theta \neq 0$ and $\mu$ approaches zero, the $Y_t$ distribution becomes symmetric, but it remains kurtotic. Higher absolute values of $\theta$ cause increased kurtosis. If $\theta \neq 0$ and $\mu > 0$, $Y_t$ has a kurtotic and right-skewed distribution, while $\mu < 0$ results in a kurtotic and left skewed distribution. Higher values of $\mu$ increase both skewness and kurtosis, but kurtosis can be scaled back by reducing $|\theta|$ (proof available from the authors).

Johnson, Kotz and Balakrishnan (pp. 34-38) indicate that both the normal and the log-normal and density are limiting cases of the $S_U$ family, which also provides for a close approximation for the Pearson family of distributions. They present the Abac for the $S_U$ family and demonstrate that for any shape factor combination below the log-normal line, there is an appropriate $S_U$ distribution. Since these shape factor results apply to the proposed expanded form of the $S_U$ family, it follows that the expanded $S_U$ family allows for any mean and variance, as well as 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.

In practice, under normality, both $\mu$ and $\theta$ would approach zero and the proposed pdf model would collapse into a normal distribution with mean $X_tB$ and variance $\sigma^2_t$ (proof available from the authors). Therefore, the null hypothesis of normality vs. the alternative of non-normality is Ho: $\theta=\mu=0$ vs. Ha: $\theta\neq 0$, $\mu\neq 0$. The null hypothesis of symmetric non-normality versus the alternative of asymmetric non-normality is Ho: $\theta\neq 0$, $\mu=0$ vs. Ha: $\theta\neq 0$, $\mu\neq 0$. The concentrated log-likelihood function that has to be maximized to estimate the non-normal pdf model defined in equation (1) is obtained using the well-known transformation technique (Mood, Graybill, and Boes):

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

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

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

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

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

The multivariate equivalent of this non-normal pdf model is obtained by assuming that each of the M random variables of interest (the potentially correlated yields from different crops in this case) follows the flexible pdf model defined in equation (1). All theoretically possible degrees of correlation among these variables are achieved by letting a multivariate normal process vector $\mathbf{V}_t \sim N(\mathbf{m},\mathbf{S})$ underlie this model, where $\mathbf{m}$ is an (M*x*1) vector of parameters and $\mathbf{S}$ is an (M*x*M) correlation matrix with unit diagonal elements and non-diagonal elements $\rho_{ij}$. The concentrated log-likelihood function that has to be maximized to estimate this multivariate non-normal pdf model is obtained using the multivariate form of the transformation technique (Mood, Graybill, and Boes):

(4) $$LL_M = \sum_{t=1}^{T} \sum_{j=1}^{M} \{ \ln(G_{jt}) - 0.5[(\mathbf{H_t}\, \mathbf{S}^{-1}).{*}\mathbf{H_t}]\} - 0.5T\ln(|\mathbf{S}|),$$

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

As suggested in recent literature, the non-random components ($X_tB$) are specified to account for the possibility of non-linear time trends in the means of the Corn Belt corn, soybean and wheat and of the West Texas Southern High and Northern Low Plains dryland cotton yield distributions. To alleviate the concerns about the ordering and power of the non-normality test (Just and

Weninger), a full model is first estimated in each case, and all statistical testing is conducted in reference to that model using the most powerful likelihood ratio tests (LRT).

The multivariate Corn Belt yield pdf model includes six parameters ($\theta_C$, $\theta_S$ and $\theta_W$, and $\mu_C$, $\mu_S$ and $\mu_W$) to account for corn {subscript (c)}, soybean (s) and wheat (w) non-normality. The West Texas cotton pdf model assumes that the degree skewness and kurtosis of the county and farm level yield distributions in both regions are the same; therefore, kurtosis and skewness are modeled by only two parameters ($\theta_{CO}$, and $\mu_{CO}$). As discussed above, the null hypothesis of normality can be tested against the alternative of non-normality by Ho:$\theta=\mu=0$ vs. Ha: $\theta\neq0$, $\mu\neq0$. Notice that since this is a joint likelihood ratio test for Ho: no kurtosis and no skewness, it does not suffer from the "double-jeopardy of normality" problem discussed in the recent literature (Just and Weninger).

In the case of Corn Belt yields, both the full ($\theta_C\neq\mu_C\neq\theta_S\neq\mu_S\neq\theta_W\neq\mu_W\neq0$) and restricted ($\theta_C=\mu_C=\theta_S=\mu_S=\theta_W=\mu_W=0$) models are multivariate. They account for any potential correlation among corn, soybean and wheat yields through the parameters $\rho_{CS}$, $\rho_{CW}$ and $\rho_{SW}$, eliminating the other potential cause of inaccuracy in the statistical significance of the non-normality tests. The mean and standard deviation of each yield distribution are estimated independently of each other, and of the distribution's skewness and kurtosis parameters, by the functions $X_{jt}B_j$ and $\sigma_{jt}$. The means of the yield distributions ($X_{jt}B_j$) are specified as third-degree polynomial functions of time:

$$(4)\qquad X_{Ct}B_C = B_{C0} + B_{C1}t + B_{C2}t^2 + B_{C3}t^3,$$

$$X_{St}B_S = B_{S0} + B_{S1}t + B_{S2}t^2 + B_{S3}t^3,$$

$$X_{Wt}B_W = B_{W0} + B_{W1}t + B_{W2}t^2 + B_{W2}t^3;$$

where t is a simple time-trend variable starting at t=1 in 1950 and ending at t=50 in 1999.

In the West Texas dryland cotton pdf model the mean ($X_{jt}B_j$) is specified as:

$$(5)\qquad X_{COt}B_{CO} = B_{00} + B_{0R}LP + B_{10}t + B_{1R}(txLP) + B_2t^2 + B_3t^3 + B_4AF_t + B_5AC_t;$$

where LP = 1 if the yield observation comes from a farm or county in the Northern Low Plains region and zero otherwise, t = 1,…29, depending on the year of the yield observation (1=1970, 29=1998), $AF_t$ = acres planted in the farm at year t in the case of a farm level yield observation and zero otherwise, $AC_t$ = acres planted in the county at year t in the case of a county level yield observation and zero otherwise. Equation (5) recognizes that the mean of the farm and county level yield distributions for a given region should be the same, but average yields could be different across regions. The latter is modeled through regional intercept and slope shifters ($B_{0R}$ and $B_{1R}$).

In the Corn Belt yield model, the standard deviation functions ($\sigma_{jt}$) are first specified as:

(6) $\sigma_{Ct} = \sigma_{C1} + \sigma_{C2}I_1 + \sigma_{C3}I_2 + \sigma_{C4}I_3 + \sigma_{C5}I_4$

$\sigma_{St} = \sigma_{S1} + \sigma_{S2}I_1 + \sigma_{S3}I_2 + \sigma_{S4}I_3 + \sigma_{S5}I_4$

$\sigma_{Wt} = \sigma_{W1} + \sigma_{W2}I_1 + \sigma_{W3}I_2 + \sigma_{W4}I_3 + \sigma_{W5}I_4$

where $I_1$=1 from 1960 through 1969 and zero otherwise, $I_2$=1 from 1970 through 1979 and zero otherwise, $I_3$=1 from 1980 through 1989 and zero otherwise, $I_4$=1 from 1990 through 1999 and zero otherwise. Thus, 1950-1959 the baseline period, and a different standard deviation is estimated for the yield distribution of every crop during each decade. A parameter and an indicator variable for the 1990-1999 decade is added to each of the variance functions when working the expanded Corn Belt yield data set. A more common heteroskedastic specification where the standard deviations are modeled by second-degree polynomial functions of time is also evaluated:

(7) $\sigma_{Ct} = \sigma_{C0} + \sigma_{C1}t + \sigma_{C2}t^2$,

$\sigma_{St} = \sigma_{S0} + \sigma_{S1}t + \sigma_{S2}t^2$,

$\sigma_{Wt} = \sigma_{W0} + \sigma_{W1}t + \sigma_{W2}t^2$;

In the West Texas dryland cotton yield pdf model the standard deviation function is:

(8) $\sigma_{COt} = \sigma_{00} + \sigma_{0R}LP + \sigma_{0L}CL + \sigma_{10}t + \sigma_{1R}(txLP) + \sigma_{1L}(txCL) + \sigma_2 t^2 + \sigma_4 AF_t + \sigma_5 AC_t$;

where CL = 1 for county level yield observations and zero otherwise, and LP, t, $AF_t$ and $AC_t$ are as defined above. This heteroskedastic specification allows for different yield variances in the initial year, which change at different rates through time, depending on the level (farms vs. county) and on the region (High vs. Low Plains). It also allows for non-linearity in the time trends of the standard deviations, and for the acres planted at the farm and county levels to affect yield variability at each of these levels.

In summary, in both cases the full models allow for yield non-normality (kurtosis and right or left skewness), third-degree polynomial time trends on the means of the yield distributions, and time-dependent heteroskedasticity. The Corn Belt yield model also permits cross-crop yield correlation. The West Texas cotton yield model estimates separate non-linear time paths for the variance at the farm and county levels, and for the two regions.

The parametric functions and parameters modeling the first four moments of the yield distributions are jointly estimated using the full information maximum likelihood procedures discussed above. This addresses the other key concern raised in recent literature: that ignoring a critical distributional characteristic (i.e., non-linearity, heteroskedasticity or multivariate correlation) when testing for another (i.e., non-normality) invalidates the result of the test.

This type of 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 non-normality, since the testing for time-trend non-linearity and heteroskedasticity without accounting for potential non-normality could affect the results of these tests. Ramirez (2000) multivariate analysis of 1909-98 U.S. corn, wheat, cotton and sorghum prices using Ramirez and Somarriba procedure to account for price autocorrelation, provides a clear example of this phenomenon.

**Results**

*Corn Belt Corn, Soybean and Wheat Yield Distributions*

The maximum likelihood parameter estimates for the full Corn Belt yield pdf models and five restricted specifications are presented in Table 1. When estimating the full model, the parameter estimates that determine the degree of non-normality in the wheat distribution ($\theta_W$ and $\mu_W$) approach zero, indicating normality. Thus, they are not reported in Table 1. LRTs for the statistical significance of the individual parameters are conducted in the case of the full and final models, by re-estimating the models with each parameter set to zero and comparing twice the difference of the maximum log-likelihood values with a $\chi^2_{(1)}$ variable. As recommended in recent literature, each of the restricted model specifications is tested against the full model.

The first of the restricted models is used to test if the means of the yield distributions follow a non-linear time-trend. A LRT statistic of $\chi^{2*}_{(6)}$ = -2[-277.477-(-266.165)] = 22.626 rejects Ho: $B_{C2}=B_{S2}=B_{W2}=B_{C3}=B_{S3}=B_{W3}=0$ in favor of Ha: at least one, $B_{C2}$, $B_{S2}$, $B_{W2}$, $B_{C3}$, $B_{S3}$, or $B_{W3} \neq 0$ at the 1% level. LRTs of Ho: $B_{C2}=B_{C3}=0$, Ho: $B_{S2}=B_{S3}=0$, and Ho: $B_{W2}=B_{W3}=0$ reject each of these hypotheses at the 5% level as well, indicating significant non-linearity in the time trends of average Corn Belt corn, soybean and wheat yields. The criticism of potential mean trend misspecification due to a priory assumption of linearity is justified.

The second restricted model is used to test for the correlation between the yield distributions. A LRT statistic of $\chi^{2*}_{(3)}$ = -2[-280.312-(-266.165)] = 28.295 strongly rejects Ho:$\rho_{CS}=\rho_{CW}=\rho_{SW}=0$ vs. Ha: at least one, $\rho_{CS}$, $\rho_{CW}$ or $\rho_{SW} \neq 0$, at the 1% level of statistical significance, indicating that at least two of the distributions are correlated. Single-parameter asymptotic Student-t tests suggest that the corn and soybean distributions are linearly correlated to each other, but not with the wheat yield distribution.

The third restricted model is used to test for heteroskedasticity. A LRT statistic of $\chi^{2*}_{(9)}$ = -2[-284.830-(-266.165)] = 37.331 rejects the null hypothesis of homoskedasticity in favor of the alternative hypothesis of heteroskedasticity, as specified in the full model, at the 1% significance level. The log-likelihood function of a jointly restricted model assuming mean linearity, non-correlation and homoskedasticity reaches a maximum value of –310.133. The LRT statistic exceeds the $\chi^2_{(18)}$ table value of 34.81 required to reject this restricted model at the 1% level of significance. Mean linearity, non-correlation and homoskedasticity are individually and jointly rejected.

The fourth restricted model specification is used to test for non-normality. As suggested in recent literature (Just and Weninger), all normality tests are conducted in relation to the full model, which includes third-degree polynomial time trends for the means, an unrestricted correlation matrix and heteroskedastic specifications. A $\chi^{2*}_{(4)}$ = -2[-275.605-(-266.165)] = 18.881 LRT statistic strongly rejects the null hypothesis of normality of both the corn and soybean yield distributions ($Ho:\theta_C=\theta_S=\mu_C=\mu_S=0$) in favor of the alternative hypothesis that at least one of the distributions is non-normal, at the 1% significance level[1].

Analogous LRTs for $Ho:\theta_C=\mu_C=0$ vs. Ha: $\theta_C\neq0$, $\mu_C\neq0$ ($\chi^{2*}_{(2)}$ = 16.275) and $Ho:\theta_S=\mu_S=0$ vs. Ha: $\theta_S\neq0$, $\mu_S\neq0$ ($\chi^{2*}_{(2)}$ = 7.450) (restricted models not presented) separately reject normality at the 2.5% level in the corn and the soybean distributions, respectively. The joint likelihood ratio tests above avoid the "double jeopardy" of other normality tests criticized in recent literature. Rejection of Ho indicates that at least one of the parameters, $\theta$ or $\mu$, is not zero at the required level of significance, which implies non-normality at that level.

The single-parameter tests (Table 1) suggest that $\mu_C$ and $\mu_S$ are individually different from zero at the 5% level, indicating that the Corn Belt corn and soybean yield distributions are skewed. The negative values of the parameter estimates for $\mu_C$ and $\mu_S$ imply left-skewness. As argued by Ramirez (1997) in more detail, the left skewness in Corn Belt corn and soybean yields is likely due to 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 log-likelihood function of a totally restricted model assuming mean linearity, non-correlation, homoskedasticity and normality reaches a maximum value of -322.75. The LRT statistic exceeds the $\chi^2_{(22)}$ table value of 37.57 required to reject this restricted model in favor of the full model at the 1% level of statistical significance. Mean linearity, homoskedasticity, non-correlation and normality are individually as well as jointly rejected.

The final model (Table 1) is formulated considering the results of the formerly discussed tests of the full vs. four restricted model specifications and of the single-parameter LRTs in the full model. It meets two essential conditions. First, neither any of the individual parameter restrictions imposed nor the set of restrictions as a whole is rejected at the 20% level of statistical significance. Second, all of the parameters included in the model are individually different from zero at the 10% level of statistical significance, according to single-coefficient LRTs.

Ramirez (1997) conclusion that annual average corn and soybean yields in the Corn Belt are heteroskedastic and non-normally distributed with a tendency towards negative (left) skewness is verified using an updated data set that includes the last ten years of Corn Belt yield data, an expanded, more refined pdf model, and addressing all of the potential model specification and statistical testing problems identified in the recent literature.

Recent literature also expresses concern about the effect of the heteroskedastic specification on the non-normality tests (Just and Weninger). Normal and non-normal yield pdf models were also estimated under the alternative, more common second-degree polynomial specifications for the standard deviation functions (the model estimation results are available from the authors upon request). A LRT rejects normality at the 2.5% significance level {$\chi^{2*}_{(4)}$ = -2[-281.2674-(-273.8772)] = 14.7803} under this alternative heteroskedastic specification as well.

The 1950-1999 Corn Belt corn and soybean yield data is plotted in Figures 1 and 2 versus the corresponding third-degree polynomial trends estimated under the full normal model. In the case of corn yields, three of the 50 observations are at least two standard deviations below the fitted curve, even when assuming a heteroskedastic process that estimates larger error-term variances for decades with increased yield volatility. No observation is two standard deviations above the fitted polynomial, and only three yield values are one standard deviation above it. A visual inspection of the corn yield data versus the normal pdf model suggests non-normality and a clear tendency towards left-skewness. In the case of soybeans, three of the 50 observations are at least two standard deviations below the fitted curve; however, the very high 1994 yield occurrence is more than two standard deviations above it. This causes a weaker rejection of non-normality than in the case of corn yields. Yet, soybean yields also appear to be left-skewed.

The 1985 corn and soybean yield distributions are simulated using the estimated normal and non-normal pdf model parameters and an adaptation of the general procedure outlined in Ramirez (1997) (details available form the authors). Figures 3 and 4 illustrate the substantial degree of left-skewness in both the corn and soybean distributions under the non-normal pdf model.

The non-normal model precisely predicts the upper limit of the corn yield distribution during the 1980s at 132 bu/acre, while the normal model implies a 23% probability of a yield occurrence above that level. The non-normal model is also accurate in predicting the probability of the six highest yields, between 120 and 132 bu/acre, observed during the 1980s (60% probability prediction, versus 25% by the normal model). In the case soybeans, the 1985 normal model forecasts a 20% likelihood of a yield occurrence above 40 bu/acre, versus a negligible probability prediction by the non-normal pdf model (Figure 4). Corn Belt soybean yields never exceeded that level during the 1980s (Figure 2). The non-normal model also provides an accurate 33% probability prediction of the three high yield occurrences of 1985, 1986 and 1987, versus 16% under the normal model. In general, for all decades, the non-normal corn and soybean pdf models are accurate in predicting the yield ceilings implied by the data, while the normal models predict between a 10% and a 30% probability of a yield occurrence above the maximum yield observed during each decade. The non-normal models are also better at predicting the probability of yields that are relatively close to the mean. Intuitively, the pdf models have to accommodate a few very low yield years with clusters of most commonly occurring yields. In doing so, the normal model forecasts a substantial proportion of improbably high yield levels and often underestimates the probability of the most commonly occurring yields. Imposing normality in these cases is inappropriate, and could substantially affect the results of any risk analysis using the simulated yield distributions.

*West Texas Dryland Cotton Yield Distributions*

The maximum likelihood parameter estimates for the full West Texas cotton yield pdf model and for six restricted specifications are presented in Table 2. As before, the statistical significance of each individual parameter is evaluated through LRTs in the case of the full and final models. The first restricted specification assumes that the kurtosis and skewness parameters ($\theta$ and $\mu$) are equal to zero, i.e. that dryland yields are normally distributed. This is used to test for non-normality. As suggested in recent literature, the normality test is conducted in relation to the full model, which includes a third degree polynomial time trend for the mean and variance of the yield distribution.

A LRT statistic of $\chi^{2*}_{(2)}$ = -2[-4721.495-(-4690.946)] = 61.098 strongly rejects the null hypothesis of normality in the West Texas dryland cotton yield distribution (Ho: $\theta_{CO}=0$, $\mu_{CO}=0$) in favor of the alternative hypothesis that the distribution is non-normal (Ho: $\theta_{CO}\neq 0$, $\mu_{CO}\neq 0$) at the 1% significance level. The single-parameter test in the full model suggests that $\mu_{CO}$ is individually different from zero at the 1% level, indicating that the West Texas dryland cotton yield distribution is skewed. The positive value of the $\mu_{CO}$ parameter estimate implies right-skewness.

The second restricted model is used to test for mean-trend non-linearity. The null hypothesis that the second and third degree polynomial trend parameters $B_2$ and $B_3$ are jointly equal to zero is rejected at the 1% significance level ($\chi^{2*}_{(2)}$ = -2[-4719.047-(-4690.946)] = 56.202). The single parameter tests in the full model (Table 2) indicate that both $B_2$ and $B_3$ are individually different from zero at the 5% level of statistical significance. The third restricted model is used to test if the mean and variance of the yield distributions are different across regions. A LRT statistic of $\chi^{2*}_{(4)}$ = -2[-4716.990-(-4690.946)] = 52.088 rejects Ho: $B_{0R}=B_{1R}=\sigma_{0R}=\sigma_{1R}=0$ vs. Ha: at least one, $B_{0R}$, $B_{1R}$, $\sigma_{0R}$ or $\sigma_{1R} \neq 0$, at the 1% significance level. The single-parameter tests in the full model identify the intercept shifter in the mean function ($B_{0R}$) and the slope shifter in the variance function ($\sigma_{1R}$) as individually significant at the 5% level.

The fourth restricted model is used to test if the variance of the yield distribution is different at the farm vs. county level. A LRT ($\chi^{2*}_{(2)}$ = -2[-4700.543-(-4690.946)] = 16.830) strongly rejects Ho: $\sigma_{0L}=\sigma_{1L}=0$ vs. Ha: at least one, $\sigma_{0L}$ or $\sigma_{1L} \neq 0$, at the 1% significance level. The single-coefficient tests in the full model, however, fail to reject the null hypothesis that either of these parameters is individually different from zero at the 10% level. Nevertheless, when $\sigma_{1L}$ is excluded from the full model, $\sigma_{0L}$ becomes statistically significant.

The fifth restricted model is used to test if the mean and variance of the farm and county level yield distributions are affected by the number of acres planted at the farm and county levels, respectively. A LRT statistic of $\chi^{2*}_{(4)}$ = -2[-4699.361-(-4690.946)] = 16.83 strongly rejects Ho: $B_{AF}=B_{AC}=\sigma_{AF}=\sigma_{AC}=0$ vs. Ha: at least one, $B_{AF}$, $B_{AC}$, $\sigma_{AF}$ or $\sigma_{AC} \neq 0$, at the 1% significance level. The single-parameter tests in the full model indicate statistically significant effects of acreage on the mean of the county level distribution and on the variance of the farm level distribution.

The full model implies that a West Texas county that plants 50,000 acres/year above the region’s average produces 17 lbs/acre (i.e. 6%) higher yields. This is consistent with the commonly held view that the counties that traditionally grow more cotton tend to have higher yields, perhaps because they are better suited to produce dryland cotton or because farmers in these counties have more widely adopted superior technologies. According to the full model, an increase in farm level area decreases yield variability, as expected. The parameter $\sigma_{0L}$ accounts for the reduction in yield variability from the small farm-level areas to the much larger county level areas. Differences in planted acres across counties do not appear to have a substantial effect on yield variability, i.e., the higher the level of aggregation, the least the effect of aggregation on yield variability. This is consistent with statistical theory.

The sixth restricted model is used to test for time-dependent heteroskedasticity. A LRT statistic of $\chi^{2*}_{(4)}$ = -2[-4696.930-(-4690.946)] = 11.968 rejects the null hypothesis of homoskedasticity with respect to time in favor of the alternative hypothesis of time-dependent heteroskedasticity at the 2.5% significance level. This means that yield variance is systematically changing through time in at least one of the regions. The single-coefficient tests indicate that $\sigma_{1R}$ is statistically different from zero but $\sigma_{10}$ is not. Yield variability has been decreasing through time in the Northern Low Plains, but it has remained constant in the Southern High Plains.

The seventh restricted model in Table 2 is the final model, formulated considering the results of the formerly discussed tests of the full vs. the other six restricted model specifications and of the single-parameter LRTs in the full model. It meets two essential conditions. First, neither any of the individual parameter restrictions imposed nor the set of restrictions as a whole is rejected at the 20% level of statistical significance. Second, all of the remaining parameters are individually different from zero at the 5% level, according to single-coefficient LRTs.

The final model implies separate dryland cotton yield distributions for the Northern High Plains and Southern Low Plains, which are kurtotic and right skewed and exhibit different variances at the farm and county levels. The 1975, 1985 and 1995 Northern High Plains distributions are simulated using the final model and an adaptation of the general procedures outlined in Ramirez

(1997). The mean, standard deviation, skewness and kurtosis coefficients of the simulated distributions (n=50,000) are calculated using standard formula (Table 3).

The pdf model predicts that dryland cotton yields reached a maximum of 267 lbs/acre in the High Plains and of 318 lbs/acre in the Low Plains, during the mid 80s, and had slightly declined to about 245 and 296 lbs/acre, respectively, by 1995. This is consistent with West Texas farmers' and researchers' beliefs that cotton yields did not increase during the last decade, and were actually lower than in the 1980s due to abnormally poor weather affecting West Texas. A concomitant factor under bad weather conditions could be the increased adoption of catastrophic crop insurance programs. Farmers covered by these programs might have less of an incentive to harvest damaged crops, thus reducing average yields per planted acre.

The model indicates that county level yield variability was initially the same in both regions. Variability has remained constant in the Southern High Plains, but it has decreased through time in the Northern Low Plains. As a result, the coefficient of variation of the Northern Low Plains yield distribution has declined substantially. The standard deviation of the yield distribution is 45 lbs/acre higher at the farm than at the county level, in both regions (Figure 5). In 1995, this would represent a 30-40% difference, depending on the region (Tables 2 and 3).

The kurtosis and skewness coefficients of the estimated yield distributions are the same by construction. The slight differences observed in Table 3 are due to the finite sample size used to simulate the distributions. Their average magnitudes (1.45 and 0.91, respectively) are substantial, and explain the noticeable right skewness of the distributions (Figures 5 to 7). Yield right skewness is also compatible with West Texas farmers and researchers intuition: Dryland cotton production systems have evolved to produce 100-500 lbs/acre (300 lbs/acre, on average), given normal rainfall conditions of 8-12 inches during the critical (May-to-August) period of the growing season. Under severe heat and very low rainfall (4-6 inches) that occurs about once a decade, many farms report very low or even zero yields. Extremely favorable temperatures and rainfall amounts of 15-20 inches occur in certain areas every 20-25 years, resulting in yields of between 600 and 750 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. In fact, the kurtosis and skewness coefficients of the 1911-1999 Lubbock, Texas, May-to-August rainfall data[2] (kurtosis=2.24, skewness=1.07) are strikingly similar to those of the simulated dryland cotton yield distribution. Including rainfall as a factor shifting the mean of the yield distribution from year to year could result on 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 full normal model {Rest. (1)} presented in Table 2 can be used to compare the simulated yield distributions that would have been obtained under the assumption of normality versus those implied by the non-normal pdf model {Rest. (7)}. Normality is rejected at the 1% level of statistical significance when comparing the full non-normal model with the full normal model. The very high LRT statistic leading to this rejection (61.098) provides strong evidence that the West Texas dryland cotton yield distribution is non-normal.

Although the non-normal model cannot be considered the true population model, it is certainly more accurate in describing yields than the normal model, and can be used to assess the potential consequences of ignoring yield non-normality. Figures 6 and 7 show the simulated 1995 farm and county level yield distributions for the Northern Low Plains, according to the normal and non-normal models. The models estimate similar means and nearly the same standard deviations at the farm level, but the differences in the implied probability distributions are substantial.

At the farm level (Figure 6) the normal model predicts a 3.6% probability of below zero yields, versus 0.4% by the non-normal pdf model (Table 3). The normal model underestimates the probability of low to moderately low yields, of between 80 and 280 lb/acre, by 19.1% (39.2% vs. 46.7%), and it overestimates the probability of moderately high-to-high yields, of between 280 and

560 lb/acre, by 12.7% (47.1% vs. 41.8%). The probability of extremely high yields, in excess of 560 lb/acre, predicted by the non-normal pdf model is 57.5% higher (6.3% vs. 4%).

The average error, obtained by aggregating the absolute values of the errors in the probability predictions within small (40 lb) intervals is 20.1%. At the county level, the normal pdf model also underestimates the probability of low to moderately low yields and overestimates the probability of moderately high to high yields, and it is particularly inaccurate in predicting the probability of very high yields (Figure 7 and Table 3). Using the normal model as an input for risk analysis would likely result in erroneous conclusions in this case as well.

**Conclusions and Recommendations**

This paper reaffirms Ramirez's (1997) findings that Corn Belt corn and soybean yields are non-normally distributed and substantially 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 (non-linear trends in the means, heteroskedasticity, kurtosis, right or left skewness and cross-distribution correlation) in a joint, full information context, which is the most efficient. The tests for non-linear trends and heteroskedasticy are conducted while allowing for any potential non-normality, and vice versa, using the additional information transmitted through the cross-distribution correlation matrix.

As recognized by the authors of previous studies, their non-rejections do not prove yield normality, since the magnitudes of the type-two errors in their normality tests are unknown. In contrast, here Corn Belt corn and soybean yields are shown to be non-normally distributed, with a small 2.5% probability of making an error in this conclusion. The consistency of the results after adding a significant amount of recent data, and under the original and an alternative heteroskedastic specification, is further evidence to the soundness of the non-normality concussions.

The case of West Texas dryland cotton yields further supports the thesis that some crop yield distributions are non-normal. The data set for this second analysis is much larger (n=850) and contains multiple observations per year. This allows for a rejection of yield normality at a very high (greater than 0.001%) level of statistical significance. This case also illustrates how to use the proposed procedures to address another issue raised in recent literature - the difference between farm and aggregate level yield variability - without having to assume yield normality. As argued above in more detail, there is no contradiction in the findings of Corn Belt corn and soybean yield distribution left-skewness and West Texas dryland cotton yield distribution right skewness. Diverse non-normality patterns can 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 crop yield distributions for policy, market, industry or farm risk analysis, or for any other purpose, should recognize that they could be non-normal, and use appropriate methods available for testing, and for estimating and simulating non-normal distributions when necessary.

**Footnotes**

[1] When assuming mean linearity, homoskedasticity and non-correlation, the null hypothesis of normality ($Ho{:}\theta_c=\theta_s=\mu_c=\mu_s=0$) is also rejected at the 1% significance level: The log-likelihood functions for the normal model reaches a maximum value of –322.75, implying a LRT statistic of $\chi^2_{(2)}$ = -2[-322.75-(-310.13)] = 25.23. The null hypothesis of normality is rejected under each of the restricted non-normal pdf model specifications presented in Table 1, as well, at the 2.5% level of statistical significance.

[2] Assuming that the mean and variance of the rainfall distribution are invariant through time, its skewness and kurtosis coefficients are estimated by standard formulae.

**References**


Anderson, J.R. "Simulation Methodology and Application in Agricultural Economics." *Rev. Market. and Agr. Econ.*42(March 1974):3-55.

Buccola, S.T. "Testing for Normality in Farm Net Returns." *Amer. J. Agr. Econ*. 68(May 1986):334-43.

Gallagher, P. "U.S. Corn Yield Capacity and Probability: Estimation and Forecasting with Non-Symmetric Disturbances." *N. C. J. Agr. Econ*. 8(January 1986):109-22.

Gallagher, P. "U.S. Soybean Yields: Estimation and Forecasting with Nonsymmetric Disturbances." *Amer. J. Agr. Econ*. 71(November 1987):796-803.

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?" *Amer. J. Agr. Econ*. 81(May 1999):287:304.

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

Moss, C.B. and J.S. Shonkwiler. "Estimating Yield Distributions with a Stochastic Trend and Non-Normal Errors." *Amer. J. Agr. Econ.*. 75(November 1993):1056-1062.

Ramirez, O.A. and E. Somarriba. "Modeling and Simulation of Autocorrelated, Non-Normal Time Series for Analyzing Risk in Diversified Agricultural Systems." *J. Agr. Resour. Econ.* 25,2(2000a): 653-668.

Ramirez, O.A. "Multivariate Transformations to Parametrically Model and Simulate Joint Price-Production Distributions under Non-Normality, Autocorrelation & Heteroskedasticity: A Tool for Assessing Risk in Agriculture." *J. Agr. App. Econ.* 32,2(August 2000b):283-297.

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.” *Amer. J. Agr. Econ.* 79(February 1997):291-205.

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.” *J. App. Stats.* 21,4(1994):289:303.

Swinton, S., and R.P. King. “Evaluating Robust Regression Techniques for Detrending Crop Yield Data with Non-Normal Errors.” *Amer. J. Agr. Econ.* 73(May 1991):446-61.

Taylor, C.R. “Two Practical Procedures for Estimating Multivariate Nonnormal Probability Density Functions.” *Amer. J. Agr. Econ.* 72(February 1990):210-217.

Table 1. Parameter estimates for six different multivariate pdf model specifications for the Corn Belt corn, soybean and wheat yield distributions.

| | Full Model | Full Normal Model[1] | Non Correlated | Homoscedastic | Linear Trends | Final Model |
|---|---|---|---|---|---|---|
| **MLV** | **-266.165** | **-275.605** | **-280.312** | **-284.830** | **-277.478** | **-269.848** |
| **LR** | | **18.881+** | **28.295+** | **37.331+** | **22.626+** | **7.366**x |
| $\theta_C$ | 1.1640** | . | 1.4083 | 0.6108 | 0.5265 | 1.2333** |
| $\mu_C$ | -8.4820** | . | -7.5364 | -9.7222 | -11.3689 | -8.3207** |
| $B_{C0}$ | 48.9749** | 45.2794 | 47.9434 | 36.5964 | 44.6256 | 49.3118** |
| $B_{C1}$ | 0.3309ns | 1.6513 | 0.3503 | 2.1671 | 1.9187 | . |
| $B_{C2}$ | 0.1037** | 0.0406 | 0.1124 | 0.0291 | . | 0.1224** |
| $B_{C3}$ | -1.6467** | -0.7848 | -1.8923 | -0.6523 | . | -1.9665** |
| $\sigma_{C1}$ | 6.4191** | 3.0001 | 8.2924 | 11.6993 | 4.0833 | 8.4648** |
| $\sigma_{C2}$ | 9.5525** | 3.1807 | 8.4697 | . | 7.1020 | 9.7336** |
| $\sigma_{C3}$ | 14.2950** | 8.4016 | 13.7969 | . | 12.2738 | 14.9422** |
| $\sigma_{C4}$ | 7.5024* | 14.1026 | 8.7211 | . | 10.2764 | 9.7336** |
| $\sigma_{C5}$ | 23.6230** | 9.4127 | 35.1803 | . | 11.1823 | 30.7209** |
| $\theta_S$ | 0.4648** | . | 0.4927 | 0.6619 | 0.4272 | 0.6628** |
| $\mu_S$ | -10.9701** | . | -10.9913 | -1.2647 | -11.1328 | -10.9786** |
| $B_{S0}$ | 22.0146** | 21.7667 | 20.9574 | 20.2606 | 20.4352 | 21.7875** |
| $B_{S1}$ | 0.1246ns | 0.1913 | 0.3669 | 0.3273 | 0.4309 | . |
| $B_{S2}$ | 0.0149** | 0.0121 | 0.0043 | 0.0076 | . | 0.0248** |
| $B_{S3}$ | -0.1943* | -0.1550 | -0.0703 | -0.1040 | . | -0.3649** |
| $\sigma_{S1}$ | 1.7167** | 1.7386 | 1.6674 | 1.7474 | 1.9984 | 2.6761** |
| $\sigma_{S2}$ | 0.5574ns | -0.0085 | 0.0661 | . | -0.0336 | . |
| $\sigma_{S3}$ | 1.1405ns | 0.8888 | 0.5536 | . | 0.6878 | . |
| $\sigma_{S4}$ | 0.9143ns | 2.0716 | 1.2093 | . | 0.7716 | . |
| $\sigma_{S5}$ | 2.0385** | 0.4836 | 2.6811 | . | 1.9791 | 3.0729** |
| $B_{W0}$ | 20.2963** | 20.1510 | 19.8743 | 19.2727 | 23.8071 | 19.9706** |
| $B_{W1}$ | 1.2332** | 1.2529 | 1.3267 | 1.3079 | 0.6362 | 1.3213** |
| $B_{W2}$ | -0.0254** | -0.0262 | -0.0300 | -0.0272 | . | -0.0302** |
| $B_{W3}$ | -0.3033* | -0.3151 | -0.3653 | -0.3144 | . | -0.3711** |
| $\sigma_{W1}$ | 3.6749** | 3.7244 | 3.6221 | 4.4717 | 4.1419 | 4.1245** |
| $\sigma_{W2}$ | -1.4983** | -1.5586 | -1.4869 | . | -1.8041 | -1.9917** |
| $\sigma_{W3}$ | 0.7501ns | 0.6987 | 0.9405 | . | 0.3266 | . |
| $\sigma_{W4}$ | 0.5691ns | 0.4709 | 0.5059 | . | -0.1597 | . |
| $\sigma_{W5}$ | 2.8706* | 2.8348 | 3.0740 | . | 2.6910 | 2.5682* |
| $\rho_{CS}$ | 0.7358** | 0.7007 | . | 0.7283 | 0.6784 | 0.7327** |
| $\rho_{CW}$ | 0.1387ns | 0.1408 | . | 0.2381 | 0.1485 | . |
| $\rho_{SW}$ | 0.1178ns | 0.1285 | . | 0.2055 | 0.1237 | . |

Notes: MLV indicates the maximum value reached by the concentrated log-likelihood function; LR indicates the likelihood ratio test statistic computed with respect to the full model; * and ** indicate that the parameter is statistically different from zero at the 10% and at the 5% level of statistical significance, respectively, according to a likelihood ratio test; ns indicates that the parameter is not different from zero at the 10% level of significance; + indicates that the restricted model is rejected at the 1% level of statistical significance; x indicates that the restricted model can not be rejected at the 10% level of statistical significance; [1] indicates the full model under the restriction of normality. The parameter estimates corresponding to $t^3$ ($B_{C3}$, $B_{S3}$ and $B_{W3}$) are multiplied by 1000.

Table 2. Parameter estimates for seven pdf models of West Texas dryland cotton yields.

| | **Full Model** | **Rest. (1) Full Normal**[1] | **Rest (2) Linear Trend** | **Rest. (3) No Reg. Diff.** | **Rest (4) No Level Effect** | **Rest. (5) No Acres Effects** | **Rest. (6) T. Homos cedastic** | **Rest. (7) Final Model** |
|---|---|---|---|---|---|---|---|---|
| **MLV** | **-4690.946** | **-4721.495** | **-4719.047** | **-4716.990** | **-4700.543** | **-4699.361** | **-4696.930** | **-4694.669** |
| **LR** | | **61.098+** | **56.202+** | **52.088+** | **19.193+** | **16.830+** | **11.968+** | **7.446**x |
| $\theta_{ct}$ | 0.2966** | . | 0.2934 | 0.2116 | 0.3012 | 0.2778 | 0.2910 | 0.2923** |
| $\mu_{ct}$ | 30.2531** | . | 29.7856 | 29.2805 | 30.1077 | 30.2835 | 29.9771 | 30.1011** |
| $B_{00}$ | 266.6783** | 265.6435 | 249.0720 | 273.6648 | 235.4011 | 252.2262 | 284.8195 | 302.1442** |
| $B_{0R}$ | 95.3106** | 92.1551 | 160.1461 | . | 92.2608 | 78.3167 | 49.8378 | 50.7793** |
| $B_{10}$ | -19.0842** | -17.1129 | -0.1901 | -11.6607 | -7.8884 | -12.4151 | -20.5551 | -23.8497** |
| $B_{1R}$ | -2.1294ns | -1.9091 | -4.7579 | . | -2.0407 | -2.0053 | 0.1089 | . |
| $B_{2}$ | 2.1444** | 1.9141 | . | 1.6097 | 1.4004 | 1.6895 | 2.0986 | 2.2932** |
| $B_{3}$ | -5.6647** | -5.0841 | . | -4.5994 | -4.2671 | -4.7648 | -5.3598 | -5.7290** |
| $B_{AF}$ | 0.0993ns | 0.1097 | -0.0128 | 0.0004 | 0.0837 | . | 0.0448 | . |
| $B_{AC}$ | 0.3420** | 0.3066 | 0.3581 | 0.0023 | 0.3158 | . | 0.2422 | 0.3131** |
| $\sigma_{00}$ | 130.9806** | 172.6902 | 173.3662 | 184.3557 | 88.6422 | 156.3592 | 187.4728 | 190.8934** |
| $\sigma_{0R}$ | 35.4857ns | 36.5598 | 68.5013 | . | 33.0098 | 23.2337 | -27.2037 | . |
| $\sigma_{0L}$ | -25.9728ns | -54.0063 | -24.5557 | -71.2007 | . | -44.2818 | -37.5010 | -45.0094** |
| $\sigma_{10}$ | 6.1280ns | 3.4662 | -2.5795 | 3.2627 | 8.6244 | 4.5327 | . | . |
| $\sigma_{1R}$ | -3.0792** | -2.6169 | -3.9351 | . | -2.7972 | -2.1434 | . | -1.4115** |
| $\sigma_{1L}$ | -0.7233ns | -0.1265 | -0.4042 | 1.4519 | . | 0.1701 | . | . |
| $\sigma_{2}$ | -0.1480ns | -0.1174 | 0.1312 | -0.1579 | -0.1928 | -0.1377 | . | . |
| $\sigma_{AF}$ | -0.0989** | -0.1102 | -0.1219 | -0.0492 | -0.0877 | . | -0.1263 | -0.1417** |
| $\sigma_{AC}$ | -0.0078ns | -0.0110 | 0.0885 | 0.0237 | -0.0241 | . | -0.0789 | . |

Notes: MLV indicates the maximum value reached by the concentrated log-likelihood function; LR indicates the likelihood ratio test statistic computed with respect to the full model; * and ** indicate that the parameter is statistically different from zero at the 10% and at the 5% level of significance, respectively, according to a likelihood ratio test; ns indicates that the parameter is not different from zero at the 10% level of statistical significance; + indicates that the restricted model is rejected at the 2.5% level of statistical significance; x indicates that the restricted model can not be rejected at the 10% level; [1] indicates the full model under the restriction of normality. The estimate corresponding to $t^3$ ($B_3$) is multiplied by 100.

Table 3. Selected statistics about the estimated/simulated Northern Low Plains and Southern High Plains 1995, 1985, and 1975 dryland cotton yield distributions.

| **Northern Low Plains** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Mean** | **Std.** | **C.V.** | **Skew** | **Kurt** | **<0** | **80-280** | **280-560** | **>560** | **A.E.** |
| **95/FA/NO** | 284.58 | 156.89 | 0.55 | 0.00 | 0.01 | 0.036 | 0.392 | 0.471 | 0.040 | 20.1 % |
| **95/FA/NN** | 295.78 | 158.68 | 0.54 | 0.91 | 1.49 | 0.004 | 0.467 | 0.418 | 0.063 | |
| | **Mean** | **Std.** | **C.V.** | **Skew** | **Kurt** | **<0** | **150-275** | **275-450** | **>450** | **A.E.** |
| **95/CO/NO** | 284.43 | 99.80 | 0.35 | 0.00 | 0.01 | 0.002 | 0.373 | 0.489 | 0.049 | 15.8 % |
| **95/CO/NN** | 295.49 | 112.78 | 0.38 | 0.91 | 1.49 | 0.000 | 0.418 | 0.424 | 0.094 | |
| **85/CO/NN** | 317.76 | 126.98 | 0.40 | 0.90 | 1.44 | -- | -- | -- | -- | -- |
| **75/CO/NN** | 283.88 | 141.37 | 0.50 | 0.91 | 1.42 | -- | -- | -- | -- | -- |
| **Southern High Plains** | | | | | | | | | | |
| | **Non-Normal** | | | | | **Normal** | | | | |
| | **Mean** | **Std.** | **C.V.** | **Skew** | **Kurt** | **Mean** | **Std.** | **C.V.** | **Skew** | **Kurt** |
| **95/FA** | 245.22 | 194.66 | 0.79 | 0.91 | 1.49 | 240.22 | 185.71 | 0.77 | 0.00 | 0.01 |
| **95/CO** | 244.94 | 148.76 | 0.61 | 0.91 | 1.49 | 240.08 | 128.62 | 0.54 | 0.00 | 0.01 |
| **85/CO** | 266.98 | 148.53 | 0.56 | 0.90 | 1.44 | -- | -- | -- | -- | -- |
| **75/CO** | 233.10 | 148.56 | 0.64 | 0.91 | 1.42 | -- | -- | -- | -- | -- |

Notes: Mean, Std., C.V., Skew and Kurt indicate the mean, standard deviation, coefficient of variation, skewness and kurtosis coefficients of the simulated yield distribution; <0 indicates the estimated probability of less than zero yields; 80-280 indicates the estimated probability of yields between 80 and 280 lbs/acre, etc.; A.E. indicates the average error in predicting yield probabilities with the normal distribution (in percentage terms); FA, CO, NO, and NN stand for farm level, county level, normal and non-normal distribution, respectively.

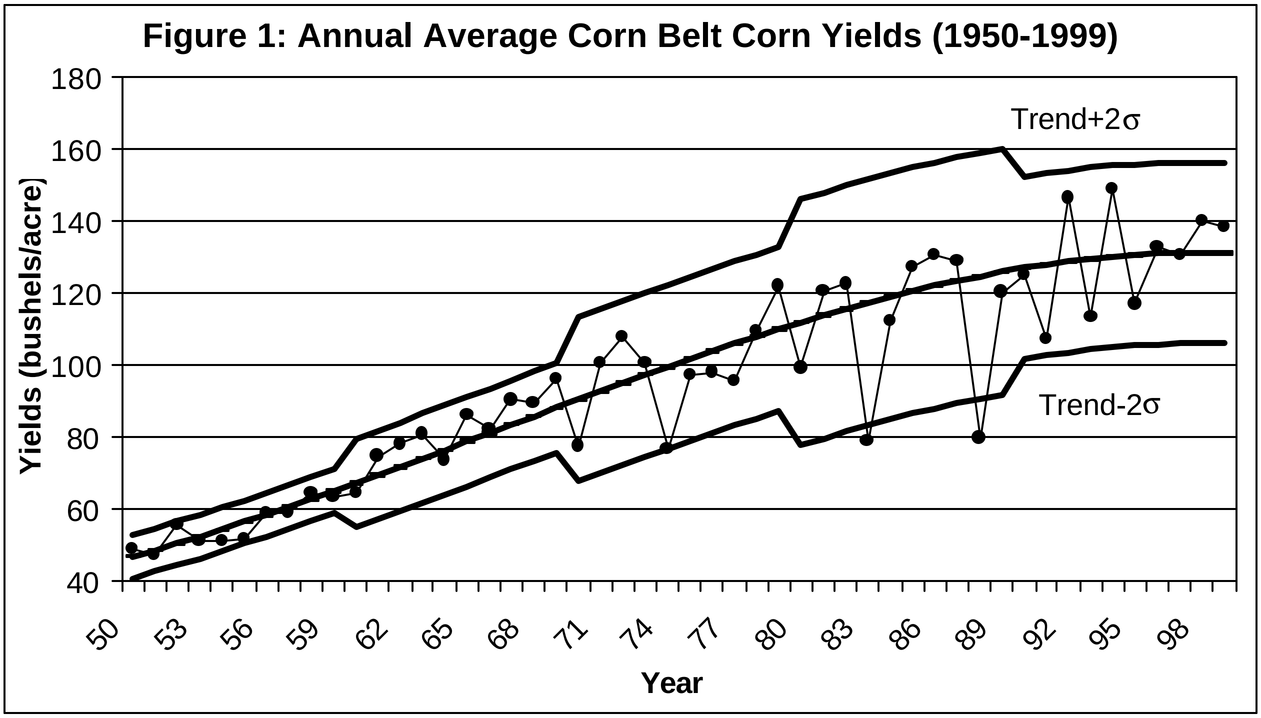


Figure 1: Annual Average Corn Belt Corn Yields (1950-1999)
Trend+2σ
Trend-2σ
Yields (bushels/acre)
Year
180
160
140
120
100
80
60
40
50
53
56
59
62
65
68
71
74
77
80
83
86
89
92
95
98

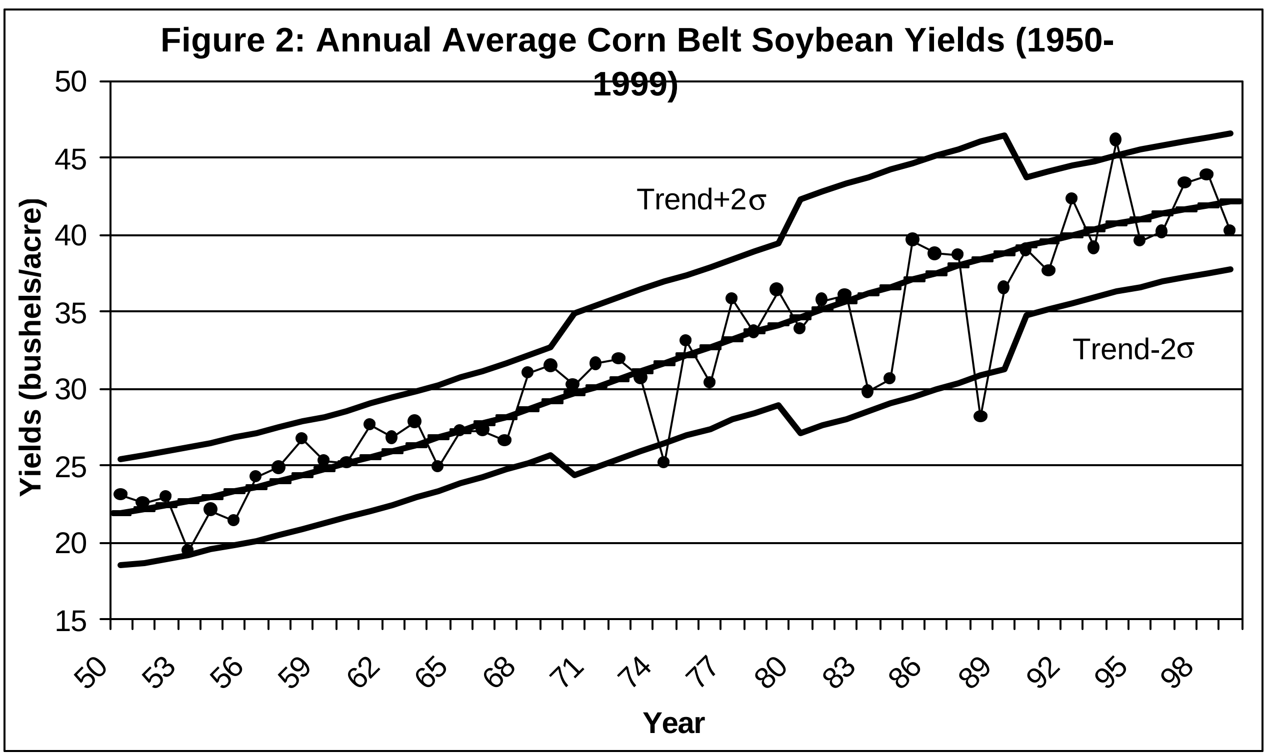


Figure 2: Annual Average Corn Belt Soybean Yields (1950-1999)
Trend+2σ
Trend-2σ
Yields (bushels/acre)
Year
50
45
40
35
30
25
20
15
50
53
56
59
62
65
68
71
74
77
80
83
86
89
92
95
98

**Figure 3: Simulated Distributions of 1985 Annual Average Corn Belt Corn Yields under Normal and Non-Normal Models**

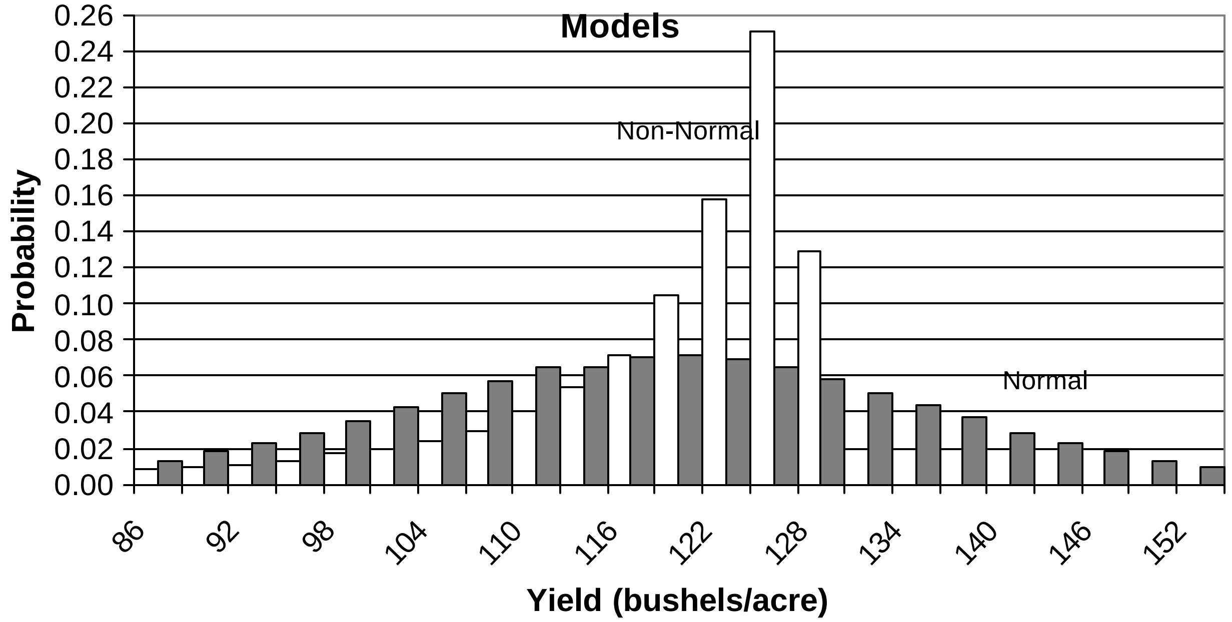
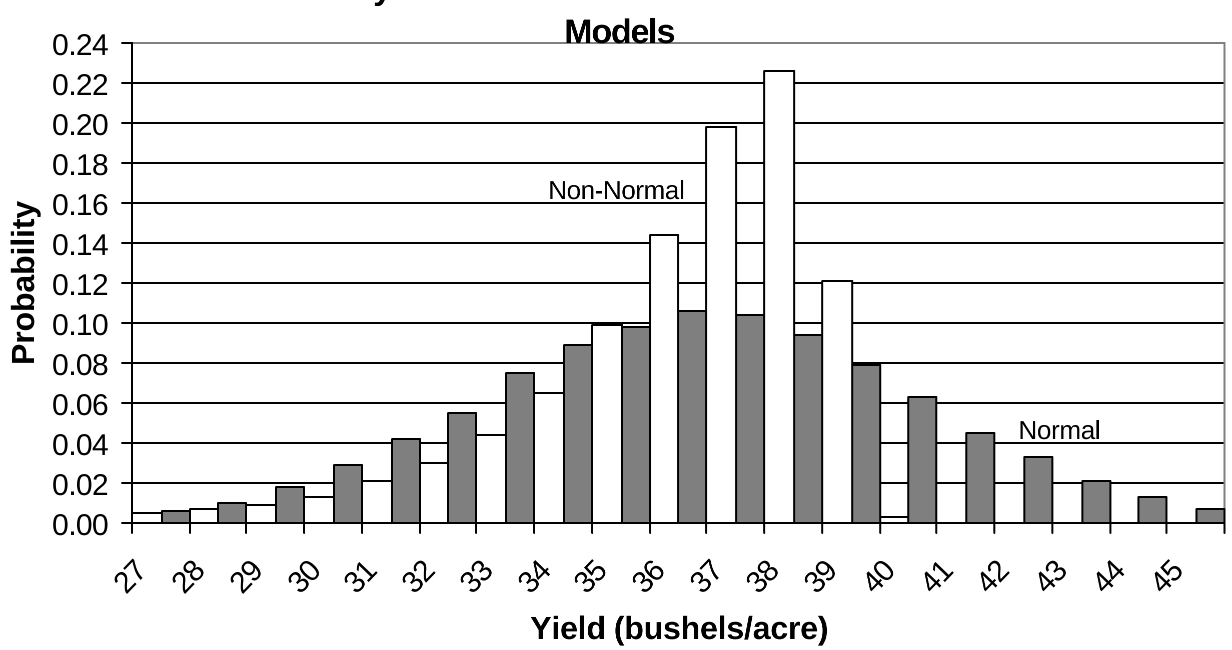



**Figure 4: Simulated Distribution of 1985 Annual Average Corn Belt Soybean Yields under Normal and Non-Normal Models**



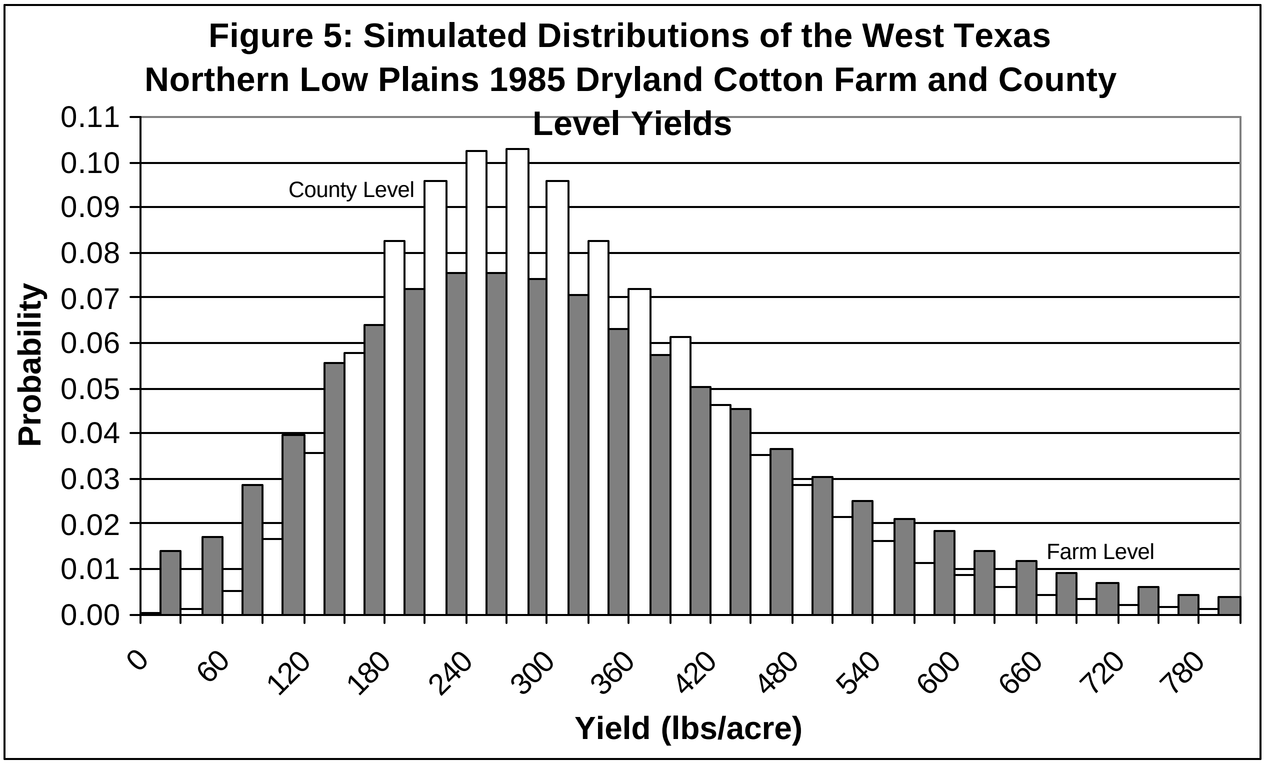
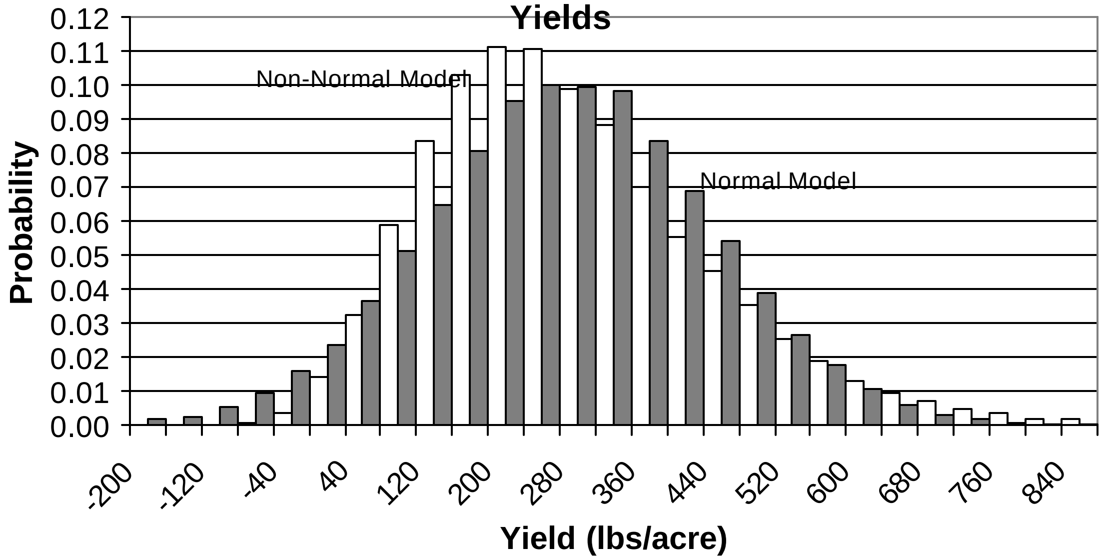


Figure 5: Simulated Distributions of the West Texas Northern Low Plains 1985 Dryland Cotton Farm and County Level Yields
Probability
0.11
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
County Level
Farm Level
0
60
120
180
240
300
360
420
480
540
600
660
720
780
Yield (lbs/acre)

**Figure 6: Simulated Distributions of the West Texas Northern Low Plains 1995 Dryland Cotton Farm Level Yields**



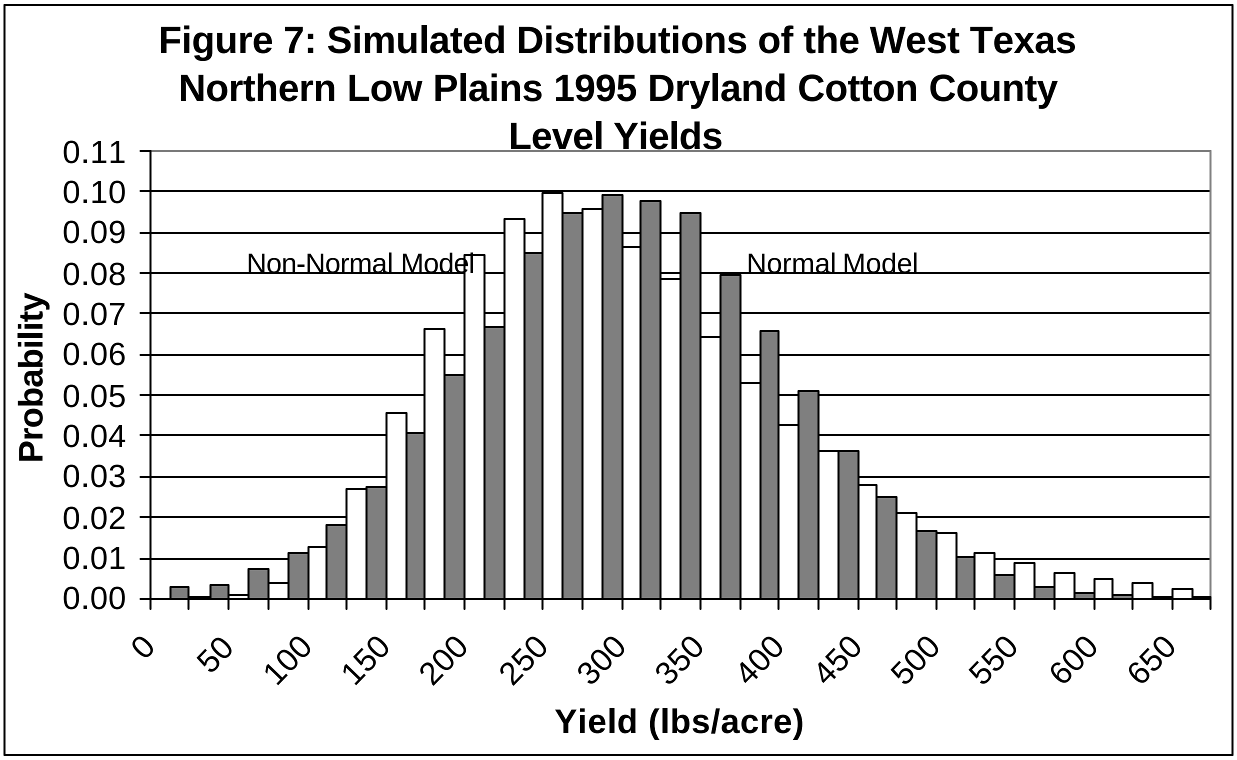


Figure 7: Simulated Distributions of the West Texas Northern Low Plains 1995 Dryland Cotton County Level Yields
Probability
Non-Normal Model
Normal Model
0.11
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
50
100
150
200
250
300
350
400
450
500
550
600
650
Yield (lbs/acre)