EVALUATING YIELD MODELS FOR CROP INSURANCE RATING

BY

CHRISTOPHER LANOUE

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Agricultural and Consumer Economics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Master's Committee:

Professor Bruce J. Sherrick, Chair
Professor Gary D. Schnitkey
Assistant Professor Nicholas D. Paulson
Assistant Professor Joshua D. Woodard

ABSTRACT

Crop insurance performance and loss rates depend directly on underlying crop yield distributions. However, there still exists much debate about how to represent the underlying crop yield distributions. Using farm-level corn and soybean yields from 1972-2008, this study examines in-sample goodness-of-fit measures of both the whole distribution and the insurance tail to compare a set of flexible parametric, semi-parametric, and non-parametric distributions in a meaningful economic context. Simulations are then conducted to investigate the out-of-sample efficiency properties of several competing distributions. The results indicate that more parameterized distributional forms fit the data better in-sample, but are generally less efficient out-of-sample – and in some cases more biased – than more parsimonious forms which also fit the data adequately, such as the Weibull. The results highlight the relative advantages of alternative distributions, in terms of the bias-efficiency tradeoff in both in- and out-of-sample frameworks.

To my family and fiancée, for their love and support.

ACKNOWLEDGMENTS

I would like to thank the people who made it possible for me to finish my thesis. First, I would like to thank my adviser, Dr. Bruce J. Sherrick, for providing me top-of-the-line resources and expert guidance. I would also like to thank Dr. Joshua D. Woodard, Dr. Gary D. Schnitkey, and Dr. Nicholas D. Paulson for their suggestions and insightful comments.

I am grateful to my family for supporting my decision to leave Fidelity Investments and enter academia in Central Illinois.

Last, but definitely not least, I am especially indebted to my wonderful fiancé for moving with me to Illinois.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This study examines alternative distributional representations of corn and soybean yields in Illinois. Two complementary approaches are used to examine both the fitting performance and the economic implications of the distributional form used. The first approach fits alternative candidate distributions to Illinois farm-level yields, and calculates goodness-of-fit measures and their corresponding implied yield insurance rates from the left tail of the distribution. The first approach shows that some distributional forms have a tendency to reproduce the in-sample variation, but are thereby less likely to predict out-of-sample yields as accurately as the simpler distributional forms. The second approach examines the out-of-sample fitting performance of alternative distributional representations under known data generating processes. The second approach demonstrates that distributions with more parameters are apt to be less precise than other distributional forms, especially in the context of insurance evaluation. The two approaches provide an in-depth examination into impacts of alternative distributional forms for modeling corn and soybean yields.

1.1 Background

An accurate understanding of yield risk is important for many different applications, including farm management decisions and risk planning. Farmers model yield risk, so that they can purchase crop insurance or use hedging strategies to offset the risk emanating from uncertain weather, demand uncertainty, and other perils. An inaccurate assessment of yield risk can cause many dollars in losses or bankruptcy to farmers whose income depends solely upon crop production. Federal crop insurance rating agencies, such as the Federal Crop Insurance Corporation (FCIC), also need to understand the underlying yield risk when issuing crop insurance products. The pricing of crop insurance not only affects program participation, but also relative indemnities paid by the FCIC and crop insurance companies. If crop insurance rates are too high, farmers may choose other methods to manage risk; conversely, rates which are too low encourage adverse selection and contribute to other problems endemic to the overprovision of insurance (e.g., land use and degradation).

Crop insurance performance depends directly on the distributions of crop yields. Despite the importance to crop insurance rating no single family of distributions or method of selection for non-parametric models is widely accepted for rating farm-level yield crop insurance. The choice of a distributional form for modeling crop yields is a complicated process because the distributional form needs to fit crop yields well in-sample, as well as accurately represent future yields. The best distributional form for modeling crop yields is a distribution that is easily implemented and is flexible enough to capture in-sample variation, but is broad enough to cover large variations in out-of-sample yields.

Traditionally, federal crop insurance premiums are set from historical loss rates and thus implicitly depend on the actual data generating process. Just and Weninger (1999) argue that aggregate time-series data cannot capture the stochastic nature of farm-level yields, and that the only way to examine the stochastic nature of yields is to examine individual farm-level yields directly. This study uses an extensive farm-level data set with more than 20 years of yields for each farm from the Illinois Farm Business Farm Management program (FBFM) from 1972-2008. The Illinois FBFM is a cooperative educational-service program that assists farmers with management decision-making and provides financial and production business analysis reports, in cooperation with the University of Illinois, Department of Agricultural and Consumer Economics. Over 2,000 farms participate in the FBFM program each year, providing dependable and long cross sectional yield histories.

1.2 Objectives

The goal of this study is to develop insights into the most accurate and efficient distributional form for modeling in-sample and out-of-sample farm-level corn and soybean yields in Illinois and ultimately identifying the economic impact on yield crop insurance rates from the distributional forms. Two approaches are used to reach this goal: (i) FBFM yields are fit to alternative distributional forms to establish how accurately the distributions fit the in-sample yields, and their accuracy in establishing yield crop insurance rates and; (ii) yields are drawn from a known data generating process to determine the out-of-sample fitting acuity and efficiency of alternative distributions when the true underlying yield model is known.

1.3 Methods

The proper distributional representation of corn and soybean yields is developed through a comparison of the fitting performance and yield crop insurance rate estimation of selected distributional forms on a large data set of Illinois farm-level yields. The large and varied farm-level yield data set in this study is paramount to advance the understanding in the crop insurance literature of a correct distributional form for modeling corn and soybean yields.

This study is arranged in seven chapters, including this introduction chapter. Chapter 2 includes a review of previous farm-level yield modeling and yield crop insurance literature, as well as how the previous literature relates to this study. Chapter 3 contains a background of the data, the detrending process, and the distributional forms utilized throughout the study. Chapter 4 is comprised of methods and results for a parametric fitting examination. The examination uses a wide and varied sample of actual farm-level yields throughout Illinois to test the fitting ability of specified parametric distributions. Chapter 5 includes two sections using the large farm-level data set and the four best fitting parametric distributional forms from the goodness-of-fit examination, in addition to a non-parametric and a semi-parametric distribution. The first section is a graphical representation of the distributional forms and the varying shapes the distributions take in fitting in-sample data. The second section estimates the yield insurance rates from each of the distributional forms and compares the rates against the farm-level empirical rates. Chapter 6 takes a known underlying data generating process – drawing from sample characteristics that are historically found on Illinois farms – and fits two parametric, one non-parametric, and one semi-parametric distribution to the data for the purpose of estimating and comparing yield insurance rates to the known process generating the samples. Chapter 7 is a compilation and discussion of the results and implications of the study.

CHAPTER 2

LITERATURE REVIEW

While there is an extensively developed body of literature on goodness-of-fit measurements among estimators, crop insurance applications depend more upon the bias and efficiency characteristics in limited regions of the distribution – generally the left tail – and thus may render standard approaches for performance evaluation less relevant. Furthermore, goodness-of-fit tests are difficult to compute for non- and semi-parametric distributions and even more difficult to compare the results to that of a parametric goodness-of-fit test. Some previous studies focus on how best to establish parametric yield distributions in rating applications, and others advance the use of non-parametric and semi-parametric methods. Moreover, past studies are principally concerned with evaluating bias in rates generated from alternative distributions relative to unknown empirical distributions. However, no studies are found that examine the efficiency of alternative distributions drawn from multiple families of distributions. The efficiency question is critically important here since typically only a few years of data are available for any farm or county in actual rating applications.

Each of the three representations: parametric, non-parametric, and semi-parametric, are discussed extensively in past crop insurance yield literature. The parametric family of distributions has the advantage of easy implementation, as well as an overall greater efficiency over the other two statistical procedures in small sample crop yield modeling. Its main disadvantage is that its parameters are not flexible enough to capture any future variability in crop yields. However, past research (Zanini, 2001; Pichon, 2002; Sherrick et al., 2004) shows that parametric distributions fit farm-level yields well when the underlying distributional form is unknown. Non-parametric distributions are more flexible than their parametric counterparts, but what they offer in flexibility they tend to lose in efficiency. Because yields used in rating crop insurance are typically estimated using small samples, the non-parametric method may be less relaible, due to out-of-sample concerns. Also, if the underlying distribution is known, a parametric distribution produces more efficient estimates than a non-parametric distribution. The non-parametric distribution in this study is a Gaussian kernel density estimator, which is similar to one of the distributions Goodwin and Ker use. The semi-parametric distribution this study examines is of the cluster-based variety, specifically a two-component mixture-of-normals. The advantage to this approach is that each yield is grouped into one of two clusters where each yield is assumed to have a similar probability distribution as the others in the cluster. This approach allows for the parameters to be estimated by an easy to implement process, and captures the variability in the sample well. Ker and Goodwin (2000) propose that yields may come from two sub-populations: years in which a catastrophic event occurs and normal weather years. If this characterization is accurate, the semi-parametric distribution is best suited to capture the variability in weather and model crop yields that come from either of the two sub-populations.

Among parametric families of distributions, many studies (Day, 1965; Ramirez, 1997; Atwood et al., 2003; Ramirez et al., 2003) reject normality as the “correct” distributional form of crop yields because of empirically prevalence of negative skewness and excess kurtosis. In contrast, Just and Weninger (1999) argue that the rejection of the normal distribution in preceding empirical research is an incorrect assumption due to methodological problems in typical yield distribution analyses. Their study identifies the common distributional problems of misspecification of the nonrandom components of yield distributions, the misreporting of statistical significance, and the use of aggregate time-series data to represent farm-level yield distributions. They imply that these problems must be solved before a rejection of normality can be confirmed. Atwood et al. state that Just and Weninger predispose their results, in favor of failing to reject for normality, by detrending at a farm level. Atwood et al. put forth a new detrending procedure of detrending individual farms by region so that the trend better follows that of the other farms in the area. The Atwood et al. study tests three detrending procedures – no trend, Just and Weninger's process of detrending by individual farm, and implicit error component detrending – using a short-term Monte Carlo simulation to generate pseudo-farms and an empirical examination on Kansas farm-level yield data. In both cases, Atwood et al. find that using implicit error component detrending by region produces more robust and powerful normality tests than detrending at an individual level. In addition, they find that detrending at an individual level tends to reduce relative insurance premia. This thesis finds that the expected value and standard deviation of the farm-level yields are sensitive to the choice of aggregation level. This thesis uses the Atwood et al. procedure by detrending each farm by its National Agricultural Statistics Service (NASS) district trend because the expected value and standard deviation sensitivity of detrending at a district level offers the greatest balance between the aggregation levels.

Other works (Nelson and Preckel, 1989; Nelson, 1990; Hennessy et al., 1997) use a conditional beta distribution to depict crop yields. The conditional beta distribution is arguably the most highly examined parametric form along with the normal distribution in empirical crop yield modeling literature. The conditional beta distribution is flexible enough to take on varying forms of skewness and kurtosis, as well as being bounded at zero and a maximum value. Nelson and Preckel use the conditional beta distribution to model the probability distribution of Iowa Agricultural Experiment Station farm-level corn yields from 1961 to 1970. Their analysis demonstrates that the conditional beta distribution is consistent with agronomic models of field crop production. Nelson further expands the examination of the conditional beta distribution by comparing crop insurance premia from both a normal and conditional beta distribution using average county yields over Iowa Agricultural Experiment Station farms for seven counties during 1964 to 1969. He provides evidence that the use of a normal distribution for crop yields consistently produces larger premium rates than when a conditional beta distribution is used. In their methodology for choosing a distribution to accurately model crop yields in rating GRP crop insurance, Skees, et al. conclude that the conditional beta distribution is best because of its thick left tail. This characteristic allows the conditional beta distribution to give a higher probability of catastrophic events occurring and hence increase the insurance premium rates to actuarial fair levels.

Still other works attempt to examine alternative parameterizations of crop yield distributions, Gallagher (1986) and Pope and Ziemer (1984) with the gamma distribution, Sherrick et al. (2004) with the Weibull distribution, and Chen and Miranda (2004) with the Burr distribution. The gamma, Weibull, and Burr distributions are similar to the conditional beta distribution and its need for relatively few parameters to capture varying degrees of skewness – positive and negative – and variances. Gallagher applies the gamma distribution to U.S. soybean yields from 1941 to 1984. His study finds that soybean yields have negative skewness stemming from the fact that "yield cannot exceed the biological potential of the plant, yet it can approach zero under blight, early frost, or extreme heat." Sherrick et al. examine whether there are economically important differences that arise from alternative parameterizations of crop yield distributions.

Their study utilizes a high quality farm-level data set from the University of Illinois Endowment Farms to compare the conditional beta, Weibull, logistic, normal, and lognormal distributions in terms of goodness-of-fit and expected payouts to APH insurance. They conclude, using the Anderson-Darling test and comparisons of the likelihood functions, the Weibull and beta distributions consistently describe the data better than the other distributions tested. In addition, they show that the distributional choice can have a significant impact in the expected value of payouts. Some previous crop yield literature, such as Chen and Miranda, use a two-parameter Burr distribution, among other distributions, to model county level crop yields in Texas. This thesis goes a step further and uses a three-parameter Burr distribution, namely the Burr XII to fit corn and soybean yields. The Burr XII distribution includes the Weibull distribution as a special case and can capture a wider range of skewness and kurtosis values than the two-parameter Burr distribution.

In addition to alternative parametric distributions, some research investigates the use of non-parametric methods (Goodwin and Ker, 1998; C.G. and Zhao, 1999; Ker and Goodwin, 2000; Norwood et al., 2004) to model crop yields. Goodwin and Ker employ a non-parametric Gaussian kernel density estimator to fit wheat and barley NASS county-level crop yields and determine fair crop insurance premium rates. The study concludes insurance premium rates estimated from a non-parametric distribution for the 1995 to 1996 Federal Group Risk crop insurance program are more actuarially accurate than current methods. Norwood et al. use a non-parametric Gaussian kernel density estimator, in addition to five other distributions from previous empirical works, to examine the in-sample and out-of-sample crop yield forecasts for 180 crop and county combinations in the U.S. Corn Belt. The in-sample forecasts are based on yields from 1967 to 1987, while the extrapolative forecasts use trends from the in-sample years to predict yields from 1988 to 1992. In this study, the forecasting rankings of the distributions are based on not only prediction error, but also how well the distribution forecasts probability statements in comparison to the observed probability statement. The non-parametric Gaussian kernel density estimator dominates the other distributional forms with the in-sample forecasts, but does not have as much power when predicting out-of-sample yields.

Still, other studies investigate the use of semi-parametric methods (Wang and Zhang, 2002; Ker and Coble, 2003). Ker and Coble apply the conditional beta, gamma, and semi-parametric kernel estimators with both a normal and conditional beta distribution to 87 Illinois counties from 1956 to 2000, to test rate efficiency. They find that the semi-parametric estimator with a normal distribution is the most efficient of those investigated. Wang and Zhang apply semi-parametric – two- and three-component mixture-of-normals – approaches to model dry land winter wheat yields for 2,945 farms with ten years of more of yields each from 1981 to 1995 in Whitman County Washington. The number of farms may be inflated since the USDA Ag Census only states there were 1,087 farms in 2002 in Whitman County. Not surprisingly, they find that the three-component mixture-of-normals distribution was the best fitting in-sample due to its ability to reflect sample variability better than a two-component mixture-of-normals.

This study further develops the work of previous empirical studies by using a unique farm-level data set to compare many competing parametric, non-parametric and semi-parametric distributional forms in terms of their ability to accurately model corn and soybean yields. Access to the long – over 35 years of historical yields – and complete – over 10,000 corn and soybean records – FBFM data set, is unprecedented in previous empirical literature.

Furthermore, this study contributes to the yield modeling literature by comparing eight widely discussed distributional forms in the same context. Most of the previous literature focuses on a single type of distribution, whereas this study examines parametric, non-parametric, and semi-parametric distributions. In comparing distributions from three distributional types, this study is able to show the advantages and disadvantages of using these distributions to model in-sample and out-of-sample corn and soybean yields.

CHAPTER 3

DATA AND DISTRIBUTIONS CONSIDERED

The suitability of distributional forms for modeling farm-level corn and soybean yields has been discussed thoroughly in past literature. This study further illuminates the subject by examining six parametric distributions: the conditional beta, the Weibull, the inverse Gaussian, the normal, the Burr XII (Singh-Madalla), and the gamma; one non-parametric distribution: the Gaussian kernel density estimator; and one semi-parametric distribution: the two-component mixture-of-normals.

3.1 Data

This study improves on previous work by utilizing a high quality, extensive farm-level dataset from the Illinois Farm Business Farm Management program (FBFM) from 1972 to 2008. FBFM, in cooperation with the University of Illinois, Department of Agricultural and Consumer Economics, is a cooperative educational-service program that assists farmers with management decision-making, and provides financial and production business analysis reports. Over 2,000 grain farms participate in the FBFM program each year, providing dependable and extensive yield histories. This dataset is unique in the United States for its long and certified corn and soybean yield data that captures a one-of-a-kind cross-section of farms.

The soybean and corn farms in the FBFM dataset contain yields from 1972 to 2008, and are from 98 of the 102 counties in Illinois. The FBFM dataset does not contain farms in Cook (Chicago), Alexander, Hardin, or Putnam counties due to combined county reporting with neighboring counties. Furthermore, in the dataset some of the yields are zeroes, which according to FBFM management are artifacts of the old data reporting system, therefore any yield that falls below 15.0 bushels/acre for corn and 6.8 bushels/acre for soybeans or the .08% percentile of all yields is removed. The initial corn filter removes 126 corn yields from 125 different farms, of which only 23 farms are used in the analysis in Chapters 4 and 5. The initial soybean filter removes 119 soybean yields from 119 different farms, of which, only 16 farms are used in the analysis in Chapters 4 and 5. There still remain many yields that appear low by modern-day production standards, but when compared to the county average are within a few standard deviations of the average. For example, the 1983 county averages for corn in Clay and Richland counties were 20 and 28 bushels/acre, respectively. However, 1983 was one of the worst drought years in Illinois history, so it is not surprising that some yields in those counties were below 20 bushels/acre. Figure 1 demonstrates the upward trend, as well as the precipitous drop in corn crop yields in 1983 and 1988 for four yield series from 1972 to 2008 – the state of Illinois, the East Southeast district, Clay county, and Clay county Farm 29. It is seen in all four of the yield series that expected yields in 1983 and 1988 are well below the historical average yields.

The farms are grouped into nine Agricultural Statistics Districts created by the National Agricultural Statistics Service (NASS). According to NASS, each district is homogenous in terms of its geography, production, and weather characteristics. The natural breaks of each district include soil type, terrain, rainfall and length of growing season. The nine Illinois National Agricultural Statistics Districts are displayed in Figure 2 – Northwest, Northeast, West, Central, East, West Southwest, East Southeast, Southwest, and Southeast.

The long current yield history and variability of the farms in the Illinois FBFM data set give the results of this study a broad audience throughout the United States Corn Belt. Many corn and soybean farms in the Corn Belt have similar characteristics to Illinois farms in this study and are able to incorporate the results into their respective yield modeling and yield crop insurance applications.

3.2 Data Screening

The FBFM data set also includes farms with short yield histories, minimal acres, and multiple year interruptions between yields. Farms with these characteristics are undesirable in this study because this study focuses on commercially viable farms with complete yield histories. A data screen is applied to the full data set in order to remove the farms with, in the following order, short yield histories, minimal acres, and multiple year interruptions. The impact of varying sample sizes is tested in Chapter 6, but smaller sample sizes are omitted at the estimation stage.

In the FBFM program, farms are given the option each year to continue their enrollment in the program for a nominal fee or opt out of the services offered by FBFM. In addition, farms may close for any number of reasons, including death of proprietor, unprofitability, or sale of land. To eliminate farms that have closed or opted out of the FBFM program without providing enough yield information to accurately conduct yield-fitting routines, a filter is implemented that only selects farms with 20 or more records. Furthermore, a filter is included that removes farms that average less than 80 acres in size from 1972-2008. This filter removes hobby farms, so that only commercially viable scale farms remain in the study. Lastly, due the nature of how FBFM collects yield data from farms, the identification numbers for farms are occasionally reused when one farm closes or opts out of FBFM participation. To ensure this study does not include overlapping farms from the same identification number, a filter is implemented that discards farms that are missing more than two consecutive years of data. This filter allows for farms that maintain one- and two-year crop rotations to remain in the study.

Application of the filters on the original yield data provides a subset of 2,088 corn farms and 1,881 soybean farms for this study. Of the corn farms, 16 have 37 or more years of data, 768 have 30 or more years of data, and 1,368 have 25 or more years of data. For the soybean farms, 13 have 37 or more years of data, 657 have 30 or more years of data, and 1,207 have 25 or more years of data. Tables 1 and 2 display the number of corn and soybeans farms passing the data screen by time period and length of yield history. The top row of the table lists the starting year of the yield history, with a maximum of 1989 since the yield samples must be 20 years or longer. The left column shows the sample length criteria and the far right column gives a sum of the number of farms that fit into each grouping. For example, there are 37 corn farms that have 26 or more years of yields with the yield history starting in 1980. This abundance of relatively large farms with long and uninterrupted sample periods is unprecedented in empirical literature and allows for a fairly large-scale evaluation of potential distributions to model crop yields.

3.3 Detrending the Yields

In order to accurately model yield distributions in the context of rating crop insurance products, the deterministic components of yields over time, namely the effects of improvements in farm technology, must be removed to allow yields from early years to be compared with yields from more recent years in a like framework. Figure 1 captures the increase in yields over time from the deterministic components. Although Figure 1 is a small cross-section of the yield series in Illinois, it is an accurate representation of the Illinois experience. The slopes from all counties and districts in the NASS database for corn and soybeans show a clear increase in yields from 1972 to 2008.

Empirical crop insurance rating papers use many different techniques to detrend data. The most common approach is to use a least squares regression with yield as the dependent variable and time as the independent variable; the coefficient on time is then identified as the trend. This method is used when the yields are stationary around a trend. Once the trend is removed the yields display constant means and standard deviations across time. A common approach to test if the yield series are stationary around a trend is to use the augmented Dickey Fuller (ADF) test. The null hypothesis of the ADF test is that the time series is non-stationary and therefore contains a unit-root, while the alternative is that the time series is stationary around a trend (Wooldridge, 2003). In 82% of the 1,881 soybean farms and 86% of the 2,088 corn farms the null hypothesis is rejected at a 1% level of significance and the alternative hypothesis of stationary around a trend is accepted. The results demonstrate that with the trend removed, the yield series becomes stationary. Therefore, the trend is removed from each yield series to control the deterministic component of the yields.

For the purpose of removing the trend, four data aggregation levels – state, district, county, and individual farm – are considered in this study. Detrending at an individual farm-level is discarded because the dispersion or standard deviation of the farm yields falls too low after detrending and negative trends start appearing in the data. There appear to be anomalies in the soybean case, as the NASS state level trend causes the standard deviation of the sample to fall below that of the district and county level. The standard deviations found by detrending at the different aggregation levels for corn remain true to form, with the dispersion ranking for the state level being the highest, followed by the district level, and the county level. The sensitivity of the standard deviation leads to the district level being chosen as the proper aggregation level for detrending because the detrended dispersion is found to be in between the individual and state levels. The district level is also selected because the weather, access to technological advances and soil qualities tend to be of the same nature for the farms within the specified district.

The trends for each district are calculated by a linear regression of the NASS district yields from 1972 to 2008, on time. The yields of each screened FBFM farm can be detrended by,

$$DetrendYield_{i,t} = OriginalYield_{i,t} + Slope_{District} * (BaseYear - Year_t)$$

where i encompasses each FBFM farm, t goes from 1972 - 2008, and $BaseYear$ is 2009.

3.4 Summary Statistics

Tables 3 and 4 provides summary statistics, farm and yield counts, and the percentage of total acreage for the Illinois FBFM filtered original and detrended yields by district, as well as aggregated by state. The summary statistics include the first four moments of the data – mean, standard deviation, skewness, and kurtosis – as well as the minimum, maximum, and coefficient of variation (CV). The CV is calculated as the standard deviation divided by the mean and allows for comparison of data sets that are centered on different values. The farm count is the number of farms that pass the data screening for each district. The yield count is the number of yields found under all of the filtered farms in the district. Finally, the percentage of total acreage by district is the sum of the average acreage for each farm in the district divided by the sum total of average acreage of all FBFM farms in the filtered sample.

3.4.1 Corn

Over 40% of the filtered corn farms, yields and percentage of total acreage are located in the Northwest and Central districts of Illinois. When the East and East Southeast districts are included they encompass approximately 70% of the total acreage and farms in the remaining sample. The Central district farms comprise of the highest detrended mean in the filtered sample at 181.8 bushels/acre. The lowest mean is found in the Southwest district, which not surprisingly contains about 5.0% of the total filtered farms and 3.4% of the total average acreage. The standard deviations range from 23.9 bushels/acres to 28.4 bushels/acres in the detrended data. Each district exhibits negative skewness or, in other terms, the mass of the yields for each district are found in the right tail of the probability distribution. For comparison, a normal distribution is symmetric and therefore has a skewness of zero, while the farm-level yield data in this study tend to be left skewed. The negative skewness of crop yields is a traditional argument explaining why normality should not be assumed when dealing with yields. The highest kurtosis values are found in the Central and Northwest districts. This means that the variance in these districts is made up of infrequent high values instead of consistent average sized deviations. Given the higher kurtosis in these districts, it is not surprising that these districts also contain the highest yields in the data set which are close to three standard deviations from the mean. The minimum detrended yield is found in the Southwest district.

3.4.2 Soybeans

For soybean farms, about 55.0% of the average acreage, yields and farms are found in the Central, East, and East Southeast districts of Illinois. Similar to the corn case, the highest average detrended yield average of 54.3 bushels/acre is in the Central district. The Central district contains the highest average of corn and soybean yields in the filtered data set, a strong indication of the high quality soil and good growing weather in the Central district. In order to compare the dispersion between corn and soybean districts on a similar level, it is necessary to analyze their respective CV values. The CV values are slightly lower in the soybean case with respect to the detrended data; therefore the soybean yields are not as spread out as the corn yields. The skewness values for the soybeans are also negative throughout all districts of Illinois and the kurtosis is approximately the same as in the corn data. The soybean yields range from a detrended minimum of 8.6 bushels/acre to a detrended maximum of 100.9 bushels/acre. The maximum is found in the Central district while the minimum is found in the Northwest, East Southeast, and Southwest districts.

3.5 Distributions Considered

The debate as to the proper distributional form for modeling crop yields in previous empirical literature leads to many differing theoretical and empirical studies, including, but not limited to, Nelson and Preckel for the use of the conditional beta; Wang and Zhang and Ker and Coble for semi-parametric distributions; Ker and Goodwin for non-parametric distributions; Sherrick et al. for the Weibull distribution; Just and Weninger for the normal distribution; Chen and Miranda for the Burr distribution; and Gallagher for the use of the gamma distribution. This study examines the following eight alternative distributional forms for modeling crop yields: the conditional beta, the normal, the gamma, the Weibull, the Burr XII, the inverse Gaussian, the Gaussian kernel density estimator, and the two-component mixture-of-normals.

3.5.1 Parametric

The conditional beta distribution is included due to the fact the upper and lower bounds give the beta more flexibility and its focus in many previous empirical research studies, (Nelson and Preckel, Nelson, and Hennessey et al.). The beta's probability density function is given by,

$$f(x, l, h|\alpha, \gamma) = \frac{(x-l)^{\alpha-1} * (h-x)^{(\gamma-1)}}{(h-l)^{(\alpha+\gamma-1)} * \mathrm{B}(\alpha, \gamma)}$$

where l is the lowest value in the range and h is the highest value in the range. Additionally, the beta distribution probability density function contains the *beta function*, given by,

$$\mathrm{B}(\alpha, \gamma) = \Gamma(\alpha) * \Gamma(\gamma) / \Gamma(\alpha + \gamma)$$

which includes the *gamma function*,

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$$

The normal distribution is included to strengthen or weaken Just and Weninger's claim that the distribution cannot easily be rejected and that it should not be excluded from crop yield modeling discussions without further inspection. Just and Weninger use multiple statistical significance tests, including the Jacque-Bera test and the R-test, to test the normality of a time series sample. Where useful this study compares several empirical distribution function and chi-square tests to test normality of farm-level yields.

The normal distribution has constant skewness and kurtosis and can possibly include values less than zero. Although these characteristics are unfavorable to modeling crop yields, as is shown in the summary statistics of each of the districts in Illinois; the economic impact in crop insurance ratings has yet to be shown. The probability density function of the normal is,

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

and its CDF is the integral of the probability density function over the range specified,

$$F(x|\mu, \sigma) = \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

The gamma distribution is able to capture many different degrees of skewness with only two parameters. It is not used as frequently as other distributional forms in empirical literature (Gallagher and Pope and Ziemer), but its ease of calculation and relatively few parameters give it credibility in this study. The gamma's probability density function is,

$$f(x|k,\theta) = x^{k-1}\frac{e^{-x/\theta}}{\theta^k\Gamma(k)}$$

and its cumulative distribution function is,

$$F(x|k,\theta) = \frac{\gamma(k, {}^{x}\!/_{\theta})}{\Gamma(k)}$$

both the PDF and the CDF contain the *gamma function*, as described above.

The Burr XII or Singh-Madalla distribution is a highly flexible parametric distribution due to its two shape parameters and one location parameter. The Burr XII distribution is capable of covering a large range of skewness and kurtosis values and has distributional forms that are easy to fit. The Burr XII's PDF is,

$$f(x|a,b,q) = a\left(\frac{q}{b}\right)\left[\left(1+\frac{x}{b}\right)^a\right]^{-(q+1)}\left(\frac{x}{b}\right)^{(a-1)}$$

and its cumulative distribution function is,

$$F(x|a,b,q) = 1 - \left[\frac{1}{\left(1+\frac{x}{b}\right)^a}\right]^q$$

The Weibull distribution is a special case of the Burr XII distribution and has many favorable characteristics for modeling crop yields. Sherrick et al. find that since the Weibull distribution is bound by zero and can cover large ranges of skewness and kurtosis values, it is a valid competing distribution to the empirically popular conditional beta distribution, in terms of goodness-of-fit and yield insurance rating. The Weibull's PDF is,

$$f(x|\alpha,\beta) = \alpha * x^{\left(\alpha - {}^{1}\!/_{\beta^\alpha}\right)} * e^{-(x/\beta)^\alpha}$$

and its CDF is,

$$F(x|\alpha,\beta) = 1 - e^{-(x/\beta)^\alpha}$$

The inverse Gaussian is utilized in insurance contexts to model positively skewed data. It is included in this thesis to support or weaken Day's argument for positive skewness in crop yields. The inverse Gaussian's PDF is given by,

$$f(x|\mu,\lambda) = \left[\frac{\lambda}{2\pi x^3}\right]^{1/2}\exp\frac{-\lambda(x-\mu)^2}{2\mu^2 x}$$

and its CDF is,

$$F(x|\mu,\lambda) = \Phi\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}-1\right)\right) + \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}+1\right)\right)$$

which contains the standard normal distribution CDF,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt$$

3.5.2 Non-Parametric

The Gaussian kernel density estimator is a highly flexible distribution that is shown by Goodwin and Ker to do an accurate job of modeling crop yields in an insurance context. The advantage offered by the kernel density estimator is that it does not have any parameters to estimate and therefore mimics the empirical distribution of a dataset. The limited ability of a non-parametric distribution to predict out-of-sample is one of its downfalls in modeling crop yields. The kernel density estimator places a bump or kernel at each yield realization and the sum of the densities is used to construct the non-parametric curve. The Gaussian kernel density approximation of its PDF is,

$$\widehat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

which contains both the *Gaussian kernel function,*

$$K\left(\frac{x - x_i}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{2h^2}}$$

and Silverman's rule of thumb for the smoothing parameter, h,

$$\widehat{h} = 0.9 * \min\left[\text{standarddeviation}, \frac{\text{interquartilerange}}{1.34}\right] * n^{-(1/5)}$$

The smoothing parameter allows for varying weights to be given to close data points in order to construct the density function. Silverman's optimal smoothing function takes away some of the density function smoothing parameter guesswork (Silverman, 1986).

3.5.3 Semi-Parametric

The two-component mixture-of-normals is a highly flexible distribution, but it loses many degrees of freedom, in terms of its six parameters, to get such flexibility. Due to more recent developments in crop insurance rating techniques, the two-component mixture-of-normals is getting more attention. The two-component mixture-of-normals captures much of the in-sample variation with its many degrees of freedom, but fails to produce accurate out-of-sample forecasts because it tends to over-fit the sample. The PDF of the two-component mixture-of-normals is,

$$f(x|\mu_1, \mu_2, \sigma_1, \sigma_2, w_1, w_2) = \sum_{i=1}^{2} w_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

and its CDF is,

$$F(x|\mu_1, \mu_2, \sigma_1, \sigma_2, w_1, w_2) = \int \sum_{i=1}^{2} w_i e^{\frac{\left[\frac{-(x-\mu_i)^2}{2\sigma_i^2}\right]}{(2\pi)^{.5}\sigma_i}} dx$$

3.6 Fitting Methods

For all of the distributions, except the Gaussian kernel density estimator, a maximum likelihood approach is used to estimate the distributional parameters. The likelihood function for a density distribution $f(x)$ with a θ parameter set over an n number of samples is defined as:

$$L = \prod_{i=1}^{n} f(X_i, \theta)$$

The parameters of the conditional beta are estimated by using the maximum likelihood estimation distribution fitting tool in @Risk for Excel 5.5. The maximum likelihood estimation in @Risk for Excel 5.5 for the conditional beta distribution uses a proprietary algorithm based on information in "Continuous Univariate Distributions" by Johnson, Katz and Balakrishnan to locate the upper and lower limits. MATLAB only has the capability to solve for the conditional beta parameters on a [0, 1] scale, therefore for the sake of consistency @Risk for Excel 5.5 is utilized.

The parameters of the remaining four parametric distributions – Weibull, gamma, Burr XII, inverse Gaussian, and normal – and the two-component mixture-of-normals are estimated by using their respective maximum likelihood estimation functions in MATLAB. MATLAB uses the Expectation Maximization (EM) algorithm to solve for the complexity of the maximum likelihood estimation of the parameters for the two-component mixture-of-normals distribution. The EM algorithm is an iterative two-step process: (i) Using estimated values for the latent variables, it computes the expectation of the log-likelihood and (ii) then it maximizes the log-likelihood found in (i) to estimate the parameters of the distribution; the parameters are next used to estimate new values for the latent variables until a maximum log-likelihood is found (Hogg et al., 2005).

The Gaussian kernel density estimator is a non-parametric distribution; therefore it does not have any parameters to estimate. The approximation of its PDF is coded into MATLAB for the purpose of producing density estimations at each crop yield data point.

CHAPTER 4

PARAMETRIC GOODNESS-OF-FIT APPLICATION

The purpose of this section is to examine the farm-level corn and soybean yield modeling capability of selected parametric distributions by comparing the distributions with several proven goodness-of-fit tests to select the top four fitting distributions. The methods section explains the intuition behind choosing the Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), and Chi-Squared (χ^2) statistical tests to compare the crop yield modeling ability of six parametric distributions - conditional beta, inverse Gaussian, gamma, normal, Burr XII, and Weibull. The results section includes the statistical test rankings of the parametric distributions over all data screened and detrended FBFM farms in Illinois and an examination of the top four fitting parametric distributions.

4.1 Methods

The K-S, A-D, and χ^2 tests each provide well-accepted methods to assess differences between the empirical CDF of a dataset and its fitted representation. To compare and rank the goodness-of-fit of each distribution, in addition to examining the ranks of the individual tests, a weighted average of the K-S, A-D, and χ^2 test ranks is created. The smaller the weighted or individual rank, the closer a distribution lies to the empirical CDF of the data and therefore a better fit.

The A-D test examines if the fit of a specified distribution is statistically significant from the empirical distribution (Stephens, 1974). The A-D test also gives more weight to the right and left tails of the data than the K-S test and is calculated as,

$$A_n^2 = n \int_{-\infty}^{+\infty} \left[F_n(x) - \widehat{F}(x)\right]^2 \Psi(x)\widehat{f}(x)dx$$

where n is the number of data points, $\Psi^2 = \frac{1}{\widehat{F}(x)\left[1-\widehat{F}(x)\right]}$, $\widehat{f}(x)$ is the fitted density function, $\widehat{F}(x)$ is the fitted cumulative distribution function, $F_n(x) = \frac{N_x}{n}$, and N_x is the number of $X_i's$ less than x.

The K-S test calculates the maximum distance between the empirical distribution function and the CDF of the distribution being analyzed (Chakravart et al., 1967). It is more sensitive to the center of the distribution than the A-D statistic and is calculated as,

$$D = \max_{1 \le i \le N} \left(F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i) \right)$$

where N is the number of data points and $F(Y_i)$ is the CDF of the continuous distribution.

The χ^2 test tends to put emphasis on both the center and the tails of the distribution and groups the data into k number of bins and finds the difference between the observed and the expected number of data points in each bin (Snedecor and Cochran, 1989). It is calculated as,

$$\mathrm{X}^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the number of observed data points in the i^{th} bin and E_i is the number of expected data points in the i^{th} bin. The number of expected data points in the i^{th} bin is expressed as,

$$E_i = N * (F(Y_u) - F(Y_l))$$

where $F(Y_u)$ is the CDF of the upper limit of the i^{th} bin of the distribution and $F(Y_l)$ is the CDF of the lower limit of the i^{th} bin of the distribution.

Due to the nature of the A-D and K-S goodness-of-fit tests, the statistics are not adjusted for different numbers of parameters. Therefore, the scope of the fitting only includes the parametric distributions with similar number of parameters. This study notes that the upper limit of the conditional beta distribution and the second shape parameter of the Burr XII distribution give slight advantages in terms of fitting in-sample, but the results from this section show this advantage for the conditional beta is miniscule or nonexistent in most cases and applications later show the problem of over-fitting to be more pronounced as a result. The mixture-of-normals and kernel density estimator are not included at this state due to different nature of parameterizations and inability to compare. Note that these more flexible distributions are predisposed to fit the actual data better than the parametric distributions [1], but they also tend to give higher probabilities to outliers than parametric distributions and often have poor out-of-sample performance. These out-of-sample issues are developed more completely in Chapter 6. The caveat to this ability to closely mirror the empirical data and give more weight to outliers is that the non- and semi-parametric distributions lack the robustness to remain accurate distributional forms for modeling farm-level crop yields in out-of-sample applications, as is shown in Chapter 7.

4.2 Results

The purpose of this section is to provide evidence as to the parametric distributions that most effectively model crop yields. To establish this evidence, the weighted and individual test rankings are summarized and the top four distributions are selected for further analyses. This section is divided into two subsections for corn and soybeans. The subsections contain goodness-of-fit results for all detrended farms passing through the acreage and yield data screening. However, 132 corn and 176 soybean farms are removed because the maximum likelihood estimation of the conditional beta distribution does not converge, due to the way the gradient is calculated for the log-likelihood estimation procedure, among other non-identifiable problems with @Risk for Excel 5.5's proprietary conditional beta estimation fitting routine. The one identifiable characteristic of the non-converging farms is that the sample size of more than 55% of the farms is between 20 and 25 years and the sample size of over 32% of the farms is between 26 and 31 years. This characteristic shows that the likelihood of the conditional beta not converging is a factor of the sample size, among other

[1] In preliminary investigations, the mixture-of-normals fit best in-sample with respect to all goodness-of-fit tests and therefore is not included going forward.

factors. The results from the remaining farms are not sensitive to dropping the observations. The comparison of goodness-of-fit rankings is examined on the remaining 1,956 corn farms and 1,705 soybean farms.

Tables 7 and 8 contain the results for this section and are laid out as follows. The far-left column contains the three goodness-of-fit tests and the weighted average of the three test ranks; the next column lists the parametric distributions. The values represent the percentage of times that a distribution is ranked between one and six; a ranking of one means that the distribution fit the best according to the specified test statistic. For example, the K-S test ranks the Burr XII distribution the best in 44.8% of the farms in the Northwest district, 37.2% second best, and 14.0% third best. The highlighted boxes indicate the distribution that fits the highest percentage of farms by ranking and goodness-of-fit test. The sums of the percentages are 100% in all cases except for the χ^2 case. The calculation of the χ^2 test allows for ties between the different parametric distributions, hence the percentages under the χ^2 rankings do not always sum to 100%.

In addition to Tables 7 and 8, summary heat-map tables of the rankings are included. The values from Tables 9 and 10 are the sum products of the rankings and the percentage of times a distribution falls into the ranking, by distributional form and district. For example, in Table 9 under the East Southeast district, the combined ranking score for the Burr XII distribution is 2.10 ($37\% \times 1 + 29\% \times 2 + 24\% \times 3 + 8\% \times 4 + 2\% \times 5 + 0\% \times 6 = 2.10$). The results are colored in such a way that the lightest colors represent the better fitting distributions and the darkest colors represent the worse fitting distributions, by district.

To test if the sample characteristics of the farms have an effect on the distributional goodness-of-fit rankings, this study also groups the farms into three bins for various sample characteristics – sample size, acreage, expected yield, and standard deviation. The results from the sample characteristic groupings are found in Tables 11-18 and are laid out as follows. The far-left column contains the three goodness-of-fit tests and the weighted average of the three test ranks; the next column has the parametric distributions being examined. The goodness-of-fit rankings are separated into three categories that correspond to the different groupings. For example, in Table 11, the rankings are separated by sample size into the following three bins – between 20 and 24 years of data, between 24 and 30 years of data, and between 30 and 37 years of data. The values represent the percentage of times that a distribution is ranked between one and six by groupings.

4.2.1 Corn

This section contains goodness-of-fit results for the 1,956 detrended FBFM corn farms across all NASS crop reporting districts. The overall results from this section are found in Table 7, while the summarized heat-map results are found in Table 9. With respect to the A-D test, the Burr XII distribution fits six of the nine districts best. The six districts in which the Burr XII performs best make up over 86% of the total FBFM reported acreage in Illinois. The Burr XII also performs best across all 1,956 farms, with 40.9% of the first place finishes. The conditional beta and Weibull are split as to the next best fitting parametric distributional form. The conditional beta comes in first place three times for the A-D test, while the Weibull does not come in first once, but does come in second in nine of the districts. Examining the A-D panel of Table 9 shows that the Weibull distribution performs slightly better than the conditional beta when all rankings are taken into consideration. The Burr XII distribution is far and away the best distribution under the A-D test; at its best the normal distribution captures, in the East Southeast district, 20.3% of the first place ranks; the gamma, in the Southeast district, 5.3% of the first place ranks; and the inverse Gaussian does not have any first place finishes across all districts. It is no surprise that the normal distribution – with its inability to capture positive or negative skewness or excess kurtosis – fits 18.1% of the time the best in the Southwest district since that district has the lowest skewness (-0.30 bushels/acre) and kurtosis (2.99 bushels/acre) values among all districts. In contrast, the normal distribution fits worst, in terms of first place rankings, in the East district, which has the highest skewness (-0.83 bushels/acre) and kurtosis (4.35 bushels/acre) among the districts. The normal distribution has the highest percentage of fourth place finishes and is consistently the next best after the conditional beta. The inverse Gaussian distribution performs poorly across all farms with almost all of its goodness-of-fit rankings falling behind the other distributional forms.

The combined score results for the K-S, χ^2, and weighted average tests are similar to the results from the A-D test with the Burr XII performing best overall. With respect to the K-S and χ^2 tests, the conditional beta has the highest percentage of first place ranks in eight of the districts, including the state total, but still lags behind the Burr XII when the combined score ranking is examined. The Weibull and normal distributions consistently come in third and fourth place, respectively, across all districts, while the gamma and inverse Gaussian distributions again fall into fifth and sixth place. From the results of all four goodness-of-fit tests, the Burr XII distribution is the overwhelming favorite for fitting in-sample farm-level corn yields. The results from the grouping by sample size, acreage, expected yield, and standard deviation do not vary from the original results of all the farms together. The Burr XII continues to fit best across all groupings and goodness-of-fit tests. This is not surprising given the extra shape parameter the Burr XII distribution has to fit data[2].

4.2.2 Soybeans

This section contains goodness-of-fit results for the 1,705 detrended FBFM soybean farms across all NASS crop reporting districts. The overall results from this section are found in Table 8, while the summarized heat-map results are found in Table 10. With respect to the A-D test, the Burr XII distribution fits the greatest number of detrended FBFM soybean farms best in all ten of the categories. The Weibull follows the Burr XII by coming in second in the nine districts and the state total. With the majority of second place finishes, the combined rank score for the Weibull is slightly lower than the score for the conditional beta. The gamma distribution fits on average 2.2% of the soybean farms best, while the inverse Gaussian is in last place across all categories. A priori, the normal distribution should have a higher percentage of first place ranks in the Southwest district and a lower percentage in the Central district, due to the each district's skewness and kurtosis values. In fact, the normal distribution fits 10.6% of the farms in the Southwest district best and only 5.9% of the farms best in the Central district. This pattern also emerges in the other districts, where higher skewness and kurtosis values lead to lower best fitting ranks for the normal distribution. This result reinforces the argument that the normal distribution is not a good candidate distribution for modeling crop yields when the yields in question are skewed or have excess kurtosis. The inverse Gaussian again performs worst which is no surprise given its inability to capture the negative skewness found in Illinois soybean farms.

The results from the K-S, χ^2, and weighted average tests contain similar results to the A-D test, with the Burr XII having the greatest number of first place rankings and the lowest combined score ranking across the districts. The rankings from the weighted average test are not surprising given the dominance the Burr XII exhibits across the other goodness-of-fit tests. The conditional beta and Weibull distributions are the next best fitting in-sample parametric distributional forms, with respect to the percentage of first

[2]If only comparing parametric distributions with two parameters, the Weibull distribution is the best fitting. This result holds for the corn farms as well as the soybean farms.

place finishes across districts and combined score ranking, followed by the normal, gamma, and inverse Gaussian. The results from the grouping by sample size, acreage, expected yield, and standard deviation do not vary from the original results of the soybean farms separated by district. The Burr XII distribution consistently outperforms all the other parametric distributions, followed in close succession by the conditional beta and Weibull distributions. The results from this section show that the Burr XII distribution is the best distributional form for modeling detrended soybean FBFM yields at a farm-level. Although, similar to the results from the corn farms, the normal distribution makes a decent case for farms that have no skewness or kurtosis.

4.3 Summary

The A-D, K-S, and χ^2 goodness-of-fit tests are used to discern the distance between the empirical CDF and the CDF of a specified distribution. These tests give approximately the same weight to differences in the tails as they do to differences in the center of the distribution. The goodness-of-fit ranking results, across districts and sample sizes, identify the Burr XII as the distributional form with the greatest percentage of first place fits across the farms. Rounding out the top four best fitting distributional forms are the conditional beta, Weibull, and normal distributions. The inverse Gaussian distribution consistently performs poorly in all cases due to its inability to capture negative skewness.

The left tail of a distribution is of primary importance in a yield-based crop insurance context and these results do not indicate which distribution fits the left tail best. For example, envision a histogram or any empirical graphical method for estimating the PDF of a detrended farm-level yield history. The left tail of the histogram is made up of years in which the yield on the farm falls below the average. If a farmer purchases yield-based crop insurance and these below average years fall beneath the yield guarantee, the insurance company would indemnify the farmer for the loss in yield. This example illustrates the importance of a further examination into how accurately a distributional form can fit the left tail of a sample of yields. The distributions must not only accurately fit in-sample data, but also be robust enough to provide accurate crop yield forecasts.

CHAPTER 5

EMPIRICAL LEFT TAIL GRAPHICAL AND INSURANCE APPLICATION

This chapter contains two separate applications quantifying the distance between the left tail of the empirical distribution and the left tail of a specified distribution. The first application takes a graphical approach to identifying the distances in the left tail, while the second application calculates the area under the left tail and compares it to the area under the empirical distribution. The purpose of this chapter is to provide additional evidence for the distribution that best represents the left tail of a farm's empirical distribution.

This chapter is divided into two main sections and each separated into three subsections – methodology, results, and a summary. The first section introduces the graphical application, while the first subsection describes the methods behind the creation of the graphs. The first results subsection contains figures of the empirical CDF from several representative corn and soybean farms overlaid with CDF plots from six distributional forms – conditional beta, normal, Weibull, Burr XII, two-component mixture-of-normals, and Gaussian kernel density – and an examination of the graphs. The second section establishes the motivation behind the empirical insurance application. The methods behind the estimation and comparison of the empirical and distributional insurance rates are put forth in the second methodology subsection. The second results subsection is made up of tables comparing the consistency and accuracy of the distributional forms for both corn and soybeans.

5.1 Graphical Comparisons

This section takes a graphical approach to identify the distributional form that fits the left tail best. The graphical approach visually demonstrates how closely the top four ranked goodness-of-fit parametric distributions – conditional beta, normal, Burr XII, and Weibull distributions –, in addition to the two-component mixture-of-normals and the kernel density estimator, fit the left tail of the farm empirical distribution. The distributions are ranked by how well they fit the empirical CDF in the left tail.

5.1.1 Graphical Methods

The graphical comparison examines the empirical CDF of each detrended farm-level yield sample from the counties of Dekalb, Mclean, and Marion overlaid with CDF graphs of parametric, semi-parametric, and non-parametric distributions. The farms from Dekalb, Mclean and Marion counties are used because they historically offer accurate representations of the Northern, Central, and Southern sections of Illinois, respectively. The clearest and most repetitive distributional patterns occurring in the farms are found within four corn farms and four soybean farms and displayed in Figures 3 and 4.

5.1.2 Graphical Results

This section contains a graphical comparison of the different distributional forms and a discussion of the various forms each distribution takes when fitting a sample of yields. This section also gives weight to the argument that the two-component mixture-of-normals and kernel density estimator tend to over-fit in-sample, while the forms of the conditional beta, Burr XII, and Weibull distributions are broad enough to capture out-of-sample variation.

5.1.2.1 Corn

The four representative detrended FBFM corn farms from Dekalb, Marion and McLean counties, found in Figure 3, contain various common distributional forms of the six distributions – Weibull, conditional beta, Burr XII, normal, kernel density estimator, and two-component mixture-of-normals. The empirical CDF plots from McLean-891 and McLean-794 show a large spike in the left tail of the sample. The kernel density estimator and the mixture-of-normals distribution fit this spike best and continue to fit the sample with great accuracy. With respect to the parametric distributions, the Weibull and Burr XII distributions capture the left tail better than the normal and conditional beta distributions. The conditional beta CDF plot over-estimates the empirical CDF in three of the four representative farms. When an outlier appears in the left tail, the Weibull and Burr XII distributions converge to the discrepancy at a faster rate than either the normal or conditional beta distributions. The normal distribution performs the worst, most likely due to its inability to capture any skewness, but when the data are almost symmetric, such as in DeKalb-316, the normal does perform almost as well as the Burr XII and Weibull distributions.

The flexibility that is exhibited by the mixture-of-normals and kernel density estimator limits their out-of-sample forecasting ability. They tend to fit the in-sample data so well that they are unable to capture any variation outside of the sample, as shown in Chapter 6. On the other hand, the Weibull and Burr XII distributions are desirable in out-of-sample crop yield forecasting because of their capability to better capture out-of-sample variation.

5.1.2.2 Soybeans

The four representative detrended FBFM soybean farms from Dekalb, Marion and McLean counties, found in Figure 4, contain various common distributional forms of the six distributions – Weibull, conditional beta, Burr XII, normal, kernel density estimator, and two-component mixture-of-normals. In all of the representative farms, the mixture-of-normals and kernel density estimator have the best fitting approach due to their flexibility in capturing the nuances of the data. The best example of this over-fitting is in McLean-926; the kernel density estimator takes many forms in almost perfectly fitting the outlier found in the left tail. This proclivity for over-fitting in-sample does not give the kernel density estimator or mixture-of-normals any room to capture out-of-sample data points. Next, the conditional beta, Burr XII, and Weibull distributions fit about the same in three out of the four representative farms. While in the fourth farm, DeKalb-189, the conditional beta distribution does not fit nearly as well as either the Burr XII or Weibull. Also, when there is an outlier in the left tail, as in McLean-788, the Burr XII and Weibull distributions converge faster to the empirical distribution and fit the empirical distribution better in the left tail than the conditional beta. Finally, when the distribution of yields is close to symmetric, such as Marion-113, the normal distribution fits as well as the other parametric distributions, but the mixture-of-normals and kernel density estimator still perform better in the left tail because they converge faster to the empirical distribution due to their flexibility.

5.1.3 Summary

The graphical examination in this section compares different distributional forms in terms of their ability to fit the empirical CDF of a sample farm. The two-component mixture-of-normals and kernel density estimator tend to over-fit the representative farms and occasionally give excess importance to one or two low production year outliers in the data. In contrast, the Burr XII, Weibull, and conditional beta distributions are able to capture most of the in-sample variation and are better suited than the two-component mixture-of-normals and kernel density estimator to accommodate out-of-sample yields from either low or high crop production years.

Although the overall fitting prowess of the distributional forms is of importance when modeling yields, when estimating yield crop insurance the left tail of the distribution is of the most importance. Estimating the left tail of the distributional form is infeasible in a graphical context due to the fitting nuances of many of the distributional forms and the sheer number of graphs that need to be examined to draw a proper conclusion. Therefore, the next section examines only the left tail of the distributions and uses a quantitative approach to compare the areas under the distributions to the area under the empirical distribution.

5.2 Empirical Insurance Comparison

This section takes an empirical insurance rating comparison approach to identify the distributional form that fits the left tail best. This empirical insurance approach compares the area under the left tails of six distributions – conditional beta, Burr XII, normal, Weibull, two-component mixture-of-normals distributions, and the kernel density estimator – with the area under the empirical distribution. The left tail in this approach is identified as the area under the distributional curve from the yield guarantee of an individual farm to zero. The yield guarantee is the coverage level multiplied by the expected detrended yield of the farm. The distributions are compared not only by how close their yield insurance rates come to the empirical yield insurance rates, but also by how consistently they converge to the empirical rates.

The empirical insurance approach is implemented by calculating the integral from zero to the yield guarantee; this area is the estimated insurance rate under the distribution. In order to capture the estimated insurance rates from the left tail of the distribution, the integral from zero to the yield guarantee is taken. The expected yield insurance rate is expressed for each farm and candidate distribution as,

$$InsRate_{i,d} = \int_{0}^{k_i} Max\{0, k_i - Y_i\} f_{d,i}(Y_i|\theta_{i,d}) \mathrm{d}Y_i$$

where $k_i = Cover \times E(Y_i)$, $E(Y_i)$ is the expected yield of the farm i, $Cover$ is the coverage level, Y_i is the yield for farm i, and $f_{d,i}(Y_i|\theta_i)$ is the probability density function of the estimated parameter set $\theta_{i,d}$ for farm i and candidate distribution d. The rate estimation approach is used throughout previous empirical literature. This study builds upon the methods and data of previous works by utilizing the comprehensive FBFM Illinois farm-level filtered data set and comparing not only the estimated rates from parametric distributions, but also the estimated rates from a non-parametric and a semi-parametric distribution. Due to the large number of farms in the state that pass the filters, the farms are aggregated up into their respective districts and a state total for closer examination.

Pope and Ziemer state that the empirical distribution function is typically the best distributional form for modeling crop yields if little or nothing is known about the underlying distribution. Since the underlying distribution is unknown for all the farms in the filtered data set, this section uses a method for calculating the rates of the empirical distribution, namely the burn rate, as the basis for comparing the estimated rates of the other distributions. The empirical burn rate in this application is the expected value of all the detrended yields for a farm that fall below the yield guarantee,

$$BurnRate_i = \frac{1}{n} \sum_{j=1}^{n} Max\{0, k - y_j\}$$

where n is the number of yields for farm i. For example, if a farm had a detrended yield history of, in bushels/acre, 113, 110, 135, 130, 180, and 170 then the average amount by which yields fall below 85% of the mean would be 7.5 bushels/acre.

5.2.1 Empirical Insurance Methods

In order to compare how different distributions fit the left tail of the empirical data, two different approaches are used. The first compares the rate estimates of a chosen distributional form to the empirical/burn rates to identify a form of bias. In this application, bias is the difference between the estimated yield insurance rate of a distributional form and the empirical/burn rate. The bias of each distributional form and farm is calculated as:

$$Bias_{i,d} = \sum_i [InsRate_{i,d} - BurnRate_i]$$

The closer the absolute bias is to zero, the better that distribution fits the sample data for this empirical insurance application.

The bias of the distributional forms has been studied extensively in previous literature, but the efficiency, or precision, of distributional forms has not been examined as extensively. This study uses the root mean square error (RMSE) to measure the precision and efficiency of the distributional fits. The RMSE of each distributional form is calculated as the square root of the sum of the differences squared between the distribution's rate estimate and the empirical/burn rate of each farm divided by the number of yields. The in-sample precision or efficiency of each distributional form is measured using the root mean squared error (RMSE) across farms as:

$$Efficiency_{i,d} = \sqrt{\frac{1}{n} \sum_i (InsRate_{i,d} - BurnRate_i)^2}$$

The purpose of this section is to examine which distributional form most accurately estimates yield crop insurance rates as it relates to the empirical/burn rates. For this comparison, the detrended yield histories of all corn and soybean farms that pass the data screening and do not have convergence errors with the conditional beta distribution are included, as is the case in Chapter 4. The six distributions – conditional beta, normal, Burr XII, Weibull, two-component mixture-of-normals, and kernel density estimator – are fit to each individual farm and the parameter estimates are used to calculate a yield crop insurance at three different coverage levels – 85%, 75%, and 65%.

Estimation of yield insurance rates is done using a computational integration method. The integration method is a variant of Gaussian quadrature, which is a numerical method to approximate the definite integral of a function. The integration numerical method is known as Gauss-Kronrod quadrature, which is a slight variation on Gaussian quadrature. This method comes pre-packaged in MATLAB.

As an example, in this study there are 154 corn farms in the Northeast district of Illinois. The expected yield of each of the individual farms is multiplied by each of the three coverage levels and saved as the yield guarantee from which to distinguish the left tail of the distribution. Next, each of the six distributional forms is fit to the individual farms and their respective parameters are estimated. Using the numerical integration method and the distributional parameter estimates, the area under each distributional curve from the yield guarantees to zero for each individual farm is calculated as the yield insurance rate. Finally, the yield insurance rates for each distributional form and coverage level are averaged over the farms in the district.

5.2.2 Empirical Insurance Results

Results are presented in two subsections for corn and soybeans. The subsections contain a comparison of yield crop insurance rates for all farms passing through the acreage and yield data screening, except for farms where the conditional beta maximum likelihood estimation did not converge. The comparison of yield insurance rates by distributional form is examined on the remaining 1,956 corn farms and 1,705 soybean farms.

This section presents results that support a conclusion as to which distribution is best for modeling in-sample crop yields, and ultimately, for evaluating yield crop insurance at the farm-level in Illinois. In addition, this section identifies distributional forms that are prone to over-fitting in-sample data by examining the efficiency results. The results are organized in Tables 19 and 20 as follows. The tables contain the comparison statistics – average, bias, and efficiency – in the far-left column and the distributional forms being compared in the next column. The bias and efficiency (RMSE) values are presented in terms of percentages relative to the empirical/ burn rate. The average portion of the table includes the empirical/burn rate, which is not included in other sections because it is implied in the calculation of the other statistics. The distributions are separated by a dashed line to distinguish the different types of distributions: parametric, parametric with three parameters, semi-parametric, and non-parametric. The results are presented by coverage level and district. The highlighted boxes within the bias and efficiency statistic demarcations indicate the distributional form that contains the lowest absolute value under each district and coverage level. For example, in the corn case of the East Southeast district the average of all the biases under a 75% coverage level for the Weibull distribution is the closest to zero when compared to the other distributions, so it is highlighted. The bold values within the bias section indicate distributional rate estimates that are larger than the empirical/burn rates. For example, in the case of corn, all the bias values under each district and coverage level for the kernel density estimator are bold because they are greater than the empirical/burn rate in every case.

5.2.2.1 Corn

The corn results are separated into FBFM farms by district and state total. The number of FBFM corn farms in each region and the total of FBFM corn farms in Illinois for this examination are: Northwest - 379, Northeast - 154, West - 88, Central - 492, East - 280, West Southwest - 140, East Southeast - 236, Southwest - 111, Southeast - 76, and state total - 1,956.

Referring to Table 19, the average of the empirical rates ranges from a high of 3.8 bushels/acre in the East district to a low of 2.3 bushels/acre in the West Southwest district, at an 85% coverage level. The averages of the empirical/burn rates for each FBFM corn farm across all districts and coverage levels are 3.1 bushels/acre, 1.3 bushels/acre, and 0.4 bushels/acre, at coverage levels of 85%, 75%, and 65%, respectively. In comparison to the empirical/burn rates – the bias statistic on the table – the conditional beta fits best in 13 of the 27 district/coverage level combinations , while the mixture-of-normals and Weibull each fit six of the combinations best and the Burr XII fits the remaining two best. The normal and kernel density estimator distributions do not produce the best fitting rates in any of the districts, in terms of bias. In absolute value terms across all the districts, the mixture-of-normals is, on average, the least biased distribution at both an 85% and a 65% coverage level, while the conditional beta is the least biased at a 75% coverage level. The rate estimates from the conditional beta and kernel density estimator distributions consistently overstate the empirical/burn rate; all 30 rates for the kernel density estimator and 25 for the conditional beta. In contrast, the rate estimates from the Weibull and mixture-of-normals distribution underestimate the empirical/burn rates 86.7% and 100% of the time, respectively. A surprising characteristic of the differences between the empirical/burn rates and the rate estimates from the distributional forms is the fact that the kernel density estimator performs the worst in terms of absolute distance from the empirical distribution rates.

In contrast, the kernel density estimator is the most efficient in-sample distribution across all districts and the state, with respect to the RMSE statistic. The kernel density estimator is the most efficient distributional form in 16 of the 27 coverage level/district combinations while the two-component mixture-of-normals is the most efficient in the remaining 11. The Weibull and Burr XII distributions have the highest RMSE values of the distributions tested with average percentages of 33.4% and 31.6%, respectively, at an 85% coverage level. The Burr XII distribution is more efficient than the Weibull in all cases because the Weibull is only one special case of the Burr XII; the Burr XII has the flexibility to capture more combinations of skewness and kurtosis than the Weibull. For all the distributions in this section, the average RMSE across all districts declines as the coverage level goes down. For instance, the RMSE value for the conditional beta is 25.0% at an 85% coverage level; 53.6% at a 75% coverage level; and 115.2% at a 65% coverage level. This result points to the fact that as the yield guarantee gets closer to zero, the distributions do not do as well in fitting the left tail. The overall corn results show that the mixture-of-normals and kernel density estimator are the best, of the distributions in this study, for measuring in-sample efficiency when the underlying distributional form is unknown.

5.2.2.2 Soybeans

The soybean results are separated into FBFM farms by district and state total. The number of FBFM soybean farms in each region and the total of FBFM soybean farms in Illinois for this examination are: Northwest - 209, Northeast - 143, West - 79, Central - 474, East - 279, West Southwest - 137, East Southeast - 224, Southwest - 94, Southeast - 66, and state total - 1,705.

Referring to Table 20, the average of the empirical rates ranges from a high of 1.0 bushel/acre in the

Southwest district to a low of 0.5 bushels/acre in the West district, at an 85% coverage level. The averages of the empirical rates for all the data screened and detrended FBFM farms across the state are, in bushels/acre, 0.8, 0.3, and 0.1, at coverage levels of 85%, 75% and 65%, respectively. Examining the bias section of Table 20, which in this application is the distance between the fitted insurance rate and the empirical/burn rate, shows that the conditional beta and the two-component mixture-of-normals distributions perform the best over all district/coverage level combinations, followed closely by the Weibull and Burr XII. The two-component mixture-of-normals fits the best in 11 of the 27 district/coverage level combinations, while the conditional beta fits best in six. The Weibull and Burr XII distributions come in third and fourth respectively, fitting the best in five and four of the district/ coverage level combinations. In contrast to the corn results, in almost all coverage level and districts, the conditional beta and Weibull distributions carry the same sign. For example, when the rate estimate of the conditional beta is less than the empirical/burn rate, the rate estimate of the Weibull is also less than the empirical/burn rate. Similar to the corn results, the kernel density estimator does not fit the best in any of the coverage level or district combinations and overstates the empirical/burn rate in 100% of the cases.

With respect to the RMSE, the two-component mixture-of-normals and kernel density estimator outperform the other distributions in all districts and coverage levels. The mixture-of-normals is the most efficient distribution in 15 of the 27 coverage level/district combinations, while the kernel density estimator is the best in the remaining 12. The parametric distributions, including the conditional beta and Burr XII, all have RMSE values between 29.5% and 39.7% at an 85% coverage level. The RMSE values of the distributions again decline as the yield guarantee goes to zero. For example, the RMSE of the normal distribution falls from a value of 29.5% at an 85% coverage level to 141.3% at a 65% coverage level. Similarly to the corn results, the two more flexible distributions – the kernel density estimator and the two-component mixture-of-normals – are more precise at fitting in-sample.

5.2.3 Summary

Overall, the non- and semi-parametric distributional forms are shown to fit in-sample yields quite well in terms of in-sample precision, while the parametric distributions perform less well. Nevertheless, in both the goodness-of-fit examination and the empirical insurance application, the conditional beta, Burr XII, and Weibull distributions outperform all other parametric distributional forms and are still capable of representing a relatively large range of skewness and kurtosis values. The normal distribution does not tend to perform as well in either the goodness-of-fit examination or the empirical insurance rating application, most likely due to its inability to capture any variation in skewness or kurtosis.

CHAPTER 6

CORN YIELD CROP INSURANCE SIMULATION APPLICATION

The purpose of this chapter is to compare the accuracy and efficiency of crop insurance rates from the empirical, conditional beta, Weibull, and two-component mixture-of-normals distributions. Since the underlying distributional form for crop yields is unknown, this examination shows the potential impact of alternative assumptions given known conditions for the data generating process. The conditional beta, Weibull, and mixture-of-normals distributions are chosen due to the extensive previous literature focused on them. The empirical distribution is included to show that it is an unbiased distributional form. Crop yields of varying sizes, means, and standard deviations are drawn from the known underlying distributional form, either Weibull or conditional beta, to create pseudo-farms. Then, four distributions are fit to the crop yields of the pseudo-farms to estimate yield crop insurance rates and to compare the rate estimates to the underlying known rates of the pseudo-farms.

The presentation is divided into the following sections. The first section contains the methodology behind the crop yield data generating process of the pseudo-farms and the fitting methods of the empirical, conditional beta, Weibull, and mixture-of-normals distributions. The second section includes an in-depth examination of the yield crop insurance rate estimates from the empirical, conditional beta, Weibull, and mixture-of-normals distributions fit to pseudo-farms that contain similar expected yields and standard deviations as the bulk of the FBFM farms in Illinois. This approach provides a method to compare the out-of-sample fitting ability of four empirically popular distributional forms.

6.1 Methods

The characteristics of the pseudo-farms for this simulation application are based on the sample statistics of the FBFM corn farms. The expected district yields from the detrended FBFM corn farms – from Table 3 – vary from a low of 136.8 bushels/acre to a high of 181.8 bushels/acre, while the standard deviations range from a low of 24.0 bushels/acre to a high of 28.4 bushels/acre. To develop a sample of farms having a broad appeal to yield crop insurance policy makers, this application uses pseudo-farms with expected yields from 160 bushels/acre to 180 bushels/acre, in multiples of 20 and standard deviations from 20 bushels/acre to 40 bushels/acre, in multiples of ten. Although the empirical fitting application in Chapter 5 does not take into account farms with less than 20 years of yields, this simulation application contains pseudo-farms with the following sample sizes: ten years, 15 years, 20 years, and 30 years. Due to the ability of the conditional beta and Weibull distributions to capture a large range of skewness and kurtosis values traditionally found in individual corn farms, the underlying distributional forms for the pseudo-farm's crop yields are generated from both distributions.

This section explains the development and methodology behind the model used to compare the accuracy and efficiency of the empirical, conditional beta, Weibull, and two-component mixture-of-normals distributions when the underlying distributional form is known. Once either the Weibull or conditional beta distribution is selected as the underlying distributional form, the yields for each combination of sample size, mean, and standard deviation are randomly drawn. The system for drawing the random yields is done by first using a modified method-of-moments[1] approach to estimate the distributional parameters from a given mean and standard deviation and second, using the estimated parameters to draw random yield samples of a specified size.

The method-of-moments approximation for the Weibull distribution is as follows. The Weibull distribution has two parameters – a scale parameter, α, and a shape parameter, β. The shape parameter, β, is estimated given the coefficient of variation, z, by

$$1/\beta = z\left(1 + (1-z)^2 \sum_{i=0}^{n} k_i z^i\right)$$

where $z = \sigma/\mu$ and the k_i and n coefficients are given below. The approximations are good for $z \leq 1.2$, where $n = 5$ and the maximum difference is $3.64e^{-6}$.

$$\begin{aligned} k_1 &= -0.001946641 \\ k_2 &= 0.153109251 \\ k_3 &= -0.083543480 \\ k_4 &= 0 \\ k_5 &= 0.007454537 \end{aligned}$$

The scale parameter, α, is estimated, given the shape parameter, β, and the mean, μ, as

$$a = [\Gamma(1 + 1/b)/\mu]^{\beta}$$

where Γ is the $gammafunction$ (Garcia, 1981). Once the parameters are estimated using method-of-moments, the estimated parameters and the sample size are used as arguments for a Weibull random number generating function in MATLAB. The sample size dictates the size of the random sample yield vector.

The method-of-moments approximation for the conditional beta is as follows. The conditional beta distribution has two shape parameters, α and β, as well as an upper and lower limit. The lower limit is bounded at zero. The two shape parameters and the function for the upper limit are expressed as,

$$\alpha = \left(\frac{\mu}{h}\right)\left(\frac{\left(\frac{\mu}{h}\right)\left(1-\left(\frac{\mu}{h}\right)\right)}{\frac{v}{h^2}} - 1\right),$$

$$\beta = \left(1-\frac{\mu}{h}\right)\left(\frac{\left(\frac{\mu}{h}\right)\left(1-\left(\frac{\mu}{h}\right)\right)}{\frac{v}{h^2}} - 1\right),$$

$$h = \mu + 3 * \sigma$$

where μ, v, and σ are the sample mean, variance, and standard deviation, respectively. There is a caveat as to how the sample yields from an underlying conditional beta distribution are drawn once the estimated parameters are approximated by method-of-moments. The random number generating function in MATLAB

[1]The process of method-of-moments equates the first two moments of the sample – mean and standard deviation – to the first two moments of the selected distribution in order to estimate the distributional parameters.

outputs values on a scale from zero to one and therefore does not take into account the upper limit. The approach this study puts forth to overcome this issue is to use the two parameter estimates of the conditional beta distribution and the sample size as arguments in the beta distribution random number generating function in MATLAB and then multiply the subsequent numbers by the upper limit to place the yields within the range of zero to the upper limit.

Given 5,000 randomly generated pseudo corn yield sets drawn from the conditional beta or Weibull distributions in groups by sample size, expected yield, and standard deviation, the yield crop insurance rates at coverage levels of 85%, 75%, and 65% are calculated for the empirical, conditional beta, Weibull, and mixture-of-normals distributions using maximum likelihood estimation fitting routines in MATLAB. MATLAB does not include a maximum likelihood estimation fitting routine for the conditional beta distribution. To work around this exclusion, MATLAB is packaged with a custom maximum likelihood estimation function, which takes a function and starting values as arguments in order to solve for the specified parameters of the function. In order to get sufficient starting values for this routine, each sample of data is scaled down by 110% of the sample maximum. The new data falls between zero and one, the pre-packaged MATLAB two-parameter beta maximum likelihood estimation fitting routine is able to return values for the two shape parameters. Using these values, in addition to 110% of the upper limit and zero, as the starting values and the conditional beta PDF function, a conditional beta MLE fitting routine is constructed in MATLAB. The crop insurance rates for the empirical distribution are calculated in the same method as the empirical/burn rate in the previous chapter.

The true yield crop insurance rates are calculated using method-of-moments to solve for the distributional parameters and then the parameters are used to integrate from zero to the yield guarantee – expected yield value multiplied by coverage level. The bias and efficiency statistics are calculated with the true yield insurance rates for an out-of-sample representation of the fitting competence of each distribution.

6.2 Results

The purpose of this section is to develop a highly detailed and diverse set of results that examines the performance of alternative parameterizations when the underlying or true distributional form is known. The results are displayed in a similar way across many tables. Each table is comprised of eight panels, under each distributional form there are four panels for each of the four sample sizes. The tables are also grouped by mean and standard deviation of the underlying distributional forms. The panels contain the comparison statistics – average, bias, and RMSE – in the far-left column and the four distributional forms being compared in the next column. The bias and RMSE values are in terms of percentages relative to the known theoretical true rate and are separated by coverage level. For instance, Table 22 includes average, bias, and efficiency statistics for 5,000 pseudo-farms drawn from both a Weibull and conditional beta distribution with ten, 15, 20, and 30 years of crop yields having an expected yield of 160 bushels/acre with a standard deviation of 30 bushels/acre. The highlighted cells represent the minimum bias or RMSE for the specified statistic and coverage level; the bold values identify the bias values where the rate estimate is greater than the true rate. The first line of each table contains the true rate for the given data generating distributional form, mean, and standard deviation. The true rates are listed at the top of the table because they are not dependent on sample size, only the first two moments of the underlying distributional form.

6.2.1 Expected Yield of 160 bushels/acre

This section contains results for the pseudo-farms having expected yields of 160 bushels/acre. The results are found in Tables 21-23. The comparison of the estimated yield crop insurance rates from the empirical, conditional beta, Weibull, and mixture-of-normals distributions is done over varying sample sizes, – ten, 15, 20, and 30; data generating processes – Weibull and conditional beta; standard deviations – 20, 30, and 40 bushels/acre; and coverage levels – 85%, 75%, and 65%.

The true rates from both the underlying distributional forms are close in value, but the rates from the underlying Weibull tend to be higher until the standard deviation of the pseudo-farms goes above 40 bushels/acre at an 85% coverage level and then the true rates from the underlying conditional beta distribution are greater. The cause of this is that the upper limit of the conditional beta is dependent on the standard deviation, therefore at higher standard deviation levels the conditional beta is stretched out to accommodate the potential of higher yields and it converges to zero at a slower rate.

With respect to the bias statistic, the fitted conditional beta distribution consistently overstates rates, ranging from 17.7%-97.3% for a small sample size of ten at the 85% coverage level, and larger at lower coverage levels. As the sample size increases the rate estimates from the conditional beta become closer to the true rate across all coverage levels. For example, at a coverage level of 85% and a sample size of 30, the bias percentage of the conditional beta ranges from 5.8%-23.4%. The conditional beta overstates the underlying true rate in 100% of the 72 combinations. Across all sample sizes at an 85% coverage level, the conditional beta overstates the true rate by approximately 25.4% more than the empirical rate. The rates from the empirical distribution are on average 2.2% lower than the Weibull rate estimates and that absolute difference becomes greater as the sample size increases. In more than 76% of the 72 cases, the Weibull distribution overstates the true rate, but at a much smaller magnitude than the conditional beta. For example, the rate estimates from the fitted Weibull distribution range from 1.2%-17.6% above the true rates for a sample size of ten at the 85% coverage level. Opposite to the Weibull, the two-component mixture-of-normals distribution understates rates in 97.2% of the standard deviation, sample size, and underlying distributional form cases. At a coverage level of 85% and a sample size of ten, the range for which the mixture-of-normals understates the rates is 5.4%-18.2%. Not surprisingly, the empirical distribution has the lowest absolute bias percentage in 66 of the 72 combinations, while the mixture-of-normals and Weibull distributions have the lowest absolute value bias percentage in the remaining six. The rate bias from the mixture-of-normals distribution is on average 30.7% lower than the rate bias from the conditional beta distribution. The bias results from an expected yield of 160 bushels/acre show that the empirical distribution is less biased than the mixture-of-normals, conditional beta, or Weibull distributions, but as the sample size increases the differences in bias between the distributional forms approaches zero.

Next, this section compares the out-of-sample efficiency results for the empirical, mixture-of-normals, conditional beta, and Weibull distributions. The RMSE of the conditional beta rate estimates is, on average, 34.8% greater than the RMSE of the Weibull rate estimates; 24.5% greater than the RMSE of the mixture-of-normals rate estimates; and 25.1% greater than the RMSE of the empirical rate estimates at an 85% coverage level. The RMSE of the conditional beta is consistently higher than the Weibull, mixture-of-normals, or empirical distributions across all sample sizes and coverage levels, but the RMSE of the conditional beta does become smaller as the sample size increases. The RMSE of the mixture-of-normals rate estimates is, on average across all sample sizes, 10.4% greater than the RMSE of the Weibull and 0.7% smaller than the empirical at an 85% coverage level; 14.8% greater and 8.9% greater at a 75% coverage level for the Weibull and empirical distributions, respectively. The empirical RMSE values range from an average of 12.9% greater than Weibull at a sample size of ten to an average of 6.8% greater at a sample size of 30. Similar to the conditional beta, the RMSE of the mixture-of-normals becomes smaller and more precise as the sample size increases. The Weibull distribution has the lowest RMSE value in 89% of the combinations, while the mixture-of-normals has the lowest RMSE in the remaining 11%. For the empirical, conditional beta, and mixture-of-normals distributions, a larger sample size corresponds to a more efficient out-of-sample fitting, but even at higher sample sizes the Weibull distribution is still a consistently more efficient distributional form for modeling out-of-sample corn yields when the yields are centered on 160 bushels/acre.

6.2.2 Expected Yield of 180 bushels/acre

This section contains results for the pseudo-farms having expected yields of 180 bushels/acre. The tabulated results are found on Tables 24-26. The comparison of the fitted yield crop insurance rates from the empirical, conditional beta, Weibull, and mixture-of-normals distributions is done over varying sample sizes, – ten, 15, 20, and 30 – data generating processes – Weibull and conditional beta – standard deviations – 20, 30, and 40 bushels/acre – and coverage levels - 85%, 75%, and 65%.

At an 85% coverage level, the true rates from each underlying distribution converge to each other as the standard deviation increases. The convergence is slower with an expected yield of 180 bushels/acre than with an expected yield of 160 bushels/acre because at a standard deviation of 40 bushels/acre the true rates from an underlying Weibull distribution are greater by 0.05. Although out of the scope of this study, given a larger standard deviation, the true theoretical rates from an underlying conditional beta distribution should become greater than the rates from an underlying Weibull distribution.

With respect to the bias statistic and similar to the previous results of expected yields of 160 bushels/acre, the conditional beta distribution consistently overstates rates – 100% of the time – ranging from 26.2% to 130.9% for a sample size of ten at the 85% coverage level, and larger at lower coverage levels. Unexpectedly, the conditional beta rates are more biased across all sample sizes and standard deviations when the pseudo-farms are drawn from a conditional beta. As the sample size increases, the rate estimates from the conditional beta become closer to the true rate across all coverage levels. For example, at a coverage level of 85% and a sample size of 30, the bias percentage of the conditional beta ranges from 7.3% to 27.8%. The Weibull distribution overstates the true rates 79.2% of the time, while the two-component mixture-of-normals understates the true rates 95.8% of the time. Again, the lowest absolute value bias percentage in 66 of the 72 combinations comes from the empirical distribution, while the remaining six are split between the Weibull and mixture-of-normals. In examining the different underlying distributional forms, the Weibull distribution tends to exhibit smaller absolute value percentages from the true rate when the underlying distributional form is Weibull, while the mixture-of-normals performs better than the Weibull in almost all cases where the underlying distributional form is conditional beta.

In comparing the out-of-sample efficiency results for the empirical, mixture-of-normals, conditional beta, and Weibull distributions the RMSE values of the conditional beta rate estimates is, on average across all sample sizes, 45.8% greater than the RMSE of the Weibull rates; 34.6% greater than the RMSE of the mixture-of-normals; and 34.7% greater than the RMSE of the empirical rates at an 85% coverage level. Similar to the previous out-of-sample efficiency results for the yields with an expected value of 160 bushels/acre, the RMSE of the conditional beta does become smaller as the sample size increases, but is not smaller than the empirical, mixture-of-normals, or Weibull distributions in any of the cases. The RMSE of the mixture-of-normals rate estimates is, on average, 11.2% greater than the RMSE of the Weibull rates at an

85% coverage level; 13.4% greater at a 75% coverage level; and 1.9% greater at a 65% coverage level. The RMSE of the empirical rate estimates is, on average, 11.1% greater than the RMSE of the Weibull at an 85% coverage level; 28.4% greater at a 75% coverage level; and 67.1% at a 65% coverage level. In 87.5% of the 72 standard deviation, sample size, and underlying distributional form combinations the Weibull distribution has the smallest RMSE value, while the mixture-of-normals is the best in the remaining 12.5%. An interesting observation is that the mixture-of-normals has the lowest RMSE value only when the underlying distributional form is the conditional beta; the standard deviation is 20 bushels/acre; and the coverage level is either 75% or 65%. In all other cases, the Weibull distribution is the most efficient or precise distributional form for modeling corn yields when the expected yield is 180 bushels/acre.

6.3 Summary

Given that the Illinois FBFM data set has data from 1972; the largest sample size possible is 37 years. An individual FBFM farm containing yields for 30 or more years is rare in Illinois and therefore out of the scope of this simulation application. The bulk of the FBFM farms in Illinois contain yields between ten and 30 years of yields. Based on this characteristic and a known underlying distributional form, the Weibull distribution is more efficient in estimating yield crop insurance rates across many standard deviations and expected corn yields than the empirical, conditional beta, or two-component mixture-of-normals distributions.

The assumption of a known underlying distributional form is unrealistic in empirical examinations given the random nature of yields in any county or district, but it is a good way to gauge the left tail out-of-sample fitting performance of one non-parametric, one semi-parametric, and two parametric distributional forms against the true theoretical left tail of an underlying distribution. The results show that the empirical distribution performs best when the bias of the fitted rates is examined, but when the focus is on the out-of-sample efficiency the Weibull distribution is more precise across all expected yields, standard deviations, coverage levels, sample sizes, and underlying distributional forms in this study. The results also show that given an underlying conditional beta distribution, the conditional beta distribution performs no better than when the underlying distributional form is Weibull. In fact the two-component mixture-of-normals comes closer to the true rate when the pseudo-farms are drawn from a conditional beta distribution.

CHAPTER 7

CONCLUSION

Issues surrounding the choice of distribution for modeling yields, as well as the manner in which one should go about evaluating and comparing them, are and remain contentious issues. This study sheds light on these issues by using a comprehensive dataset from the Illinois FBFM of commercial scale corn and soybean yields from 1972-2008 to examine alternative distributional forms for modeling crop yields and to inspect their economic implications to crop insurance rates. This chapter is divided into three sections. The first section provides a summary of the methods and findings. The second section gives examples of further research that come from the findings of this study and the implications the findings of this study provide. The third section contains the conclusions of the study.

7.1 Summary

Chapter 4 ranks the in-sample fitting performance of six parametric distributional forms – conditional beta, gamma, inverse Gaussian, normal, Beta XII, and Weibull. The distributional representations are fit using maximum likelihood estimation to Illinois detrended FBFM farm-level corn and soybean yields. Three common goodness-of-fit tests – Anderson-Darling, Kolmogorov-Smirnov, and Chi-Squared – in addition to a weighted average of the three tests are used to rank the in-sample fitting of the parametric distributions. The results from this chapter show that the Burr XII distribution fits the in-sample data much better than the other parametric distributional forms. While the Weibull, conditional beta, and normal distributions fall into line behind the Burr XII.

Chapter 5 is comprised of two parts. In the first part, the top four parametric distributions from Chapter 4 are compared against a two-component mixture-of-normals distribution and a kernel density estimator in a graphical context. The distributions are fit to the detrended FBFM farms and eight representative corn and soybean farms are chosen, from three representative counties in different geographical sections of Illinois, to give a context for how the distributional forms fit in-sample. The second part takes the six distributions from part one and compares the left tail of each of the distributional forms against the empirical/burn rate. To compare the left tails, the parameters of the fitted distributions are used to estimate insurance rates. The distribution rate estimates are compared on accuracy and efficiency to the empirical distribution. The results from the two approaches show that the two-component mixture-of-normals and kernel density estimator are the most efficient distributions for modeling in-sample farm-level yields.

Chapter 6 compares the out-of-sample left-hand fitting ability of the empirical, conditional beta, Weibull, and two-component mixture-of-normals distributions when the underlying distributional form of the sample is known. Pseudo-farm yields are generated from either an underlying Weibull or conditional beta distribution having similar means, standard deviations, and sample sizes to what is traditionally found on farms in Illinois.

The empirical, conditional beta, Weibull, and mixture-of-normals distributions are fit to the pseudo-farm yields and insurance rates are estimated at coverage levels of 65%, 75%, and 85%. The rate estimates for each distribution are compared by their proximity to the underlying true theoretical rates. The results show that while the empirical distribution has lower absolute bias values, the more parsimonious Weibull distribution outperforms the empirical, conditional beta, and mixture-of-normals on the basis of out-of-sample efficiency, especially in smaller sample sizes.

7.2 Implications and Suggestions for Future Research

The scope of this study includes few farms with greater than 30 years of yields due to the small number of actual farms that contain near-perfect yield histories in the data. Nevertheless, insurers typically group large numbers of farms together with like characteristics when making rates. Thus, further research is needed in order to assess the sampling distribution questions addressed here in more realistic and comprehensive frameworks when several like risk farms are combined to estimate rates. Also, the out-of-sample analysis is based on simulated pseudo-data from known and restrictive parametric distributions, and thus the out-of-sample results found here may not always carry over to cases representing actual data for any particular application (e.g., if the data has larger tails than the fitted conditional beta and Weibull distributions in use here). Thus, frameworks need to be developed which can effectively assess out-of-sample rate performance using actual yield data.

This study includes a large and diverse data set containing individual FBFM farms with yield histories ranging from 20 to 37 years of yields. The yields from the individual FBFM farms encompass a vast range of means, standard deviations, skewness, and kurtosis values for corn and soybean farms in Illinois. The nine NASS crop reporting districts in Illinois also vary in their soil quality and expected production, among other characteristics. For example, the Southern districts tend to yield fewer bushels per acre than the central part of the state. This wide array of farms and yield histories gives the study appeal to insurance agencies and corn and soybean farmers across the United States Corn Belt. The implications of using such a diverse sample of high quality farms is that the results are overarching and are not biased by focusing on only one similar sample of farms.

This study also examines multiple families of distributional forms that are commonly included in empirical yield modeling and crop insurance research papers. This study finds it is important to test the distributions on similar grounds and not give preference to one distributional form over another. The implications of using many different types of distributions – non-parametric, semi-parametric, and parametric – in similar situations are that this study provides a unique examination into the over-fitting tendencies of non- and semi-parametric distributions comparing them to the broad distributional coverage offered by parametric distributions.

7.3 Conclusions

This study takes into consideration the previous empirical yield modeling and crop insurance literature and then attempts to add to the discussion with a broad spectrum of results and empirical methods both in-sample and out-of-sample. This study adds to the previous empirical literature by examining many alternative distributional forms and fitting them in-sample to a high quality farm-level data set to examine the economic implications of rating yield crop insurance products. This study also examines the out-of-sample accuracy and efficiency of one non-parametric, one semi-parametric, and two parametric distributional forms by drawing pseudo-farms from known underlying distributional forms.

The overall results of this study point to the Weibull distribution as the most efficient distribution out-of-sample and the kernel density estimator and mixture-of-normals in-sample. In-sample, the Burr XII distribution is the best fitting and most efficient parametric distributional form. These results are somewhat in contrast to the findings of Norwood, Roberts, and Lusk (2004), who find that the mixture-of-normals is superior to other distributions in that study, and calls into question generalization of the "best" distribution (whether in-sample or out-of-sample) for any particular application. The results of the simulations illustrate the bias-efficiency tradeoff when evaluating distributions with different levels of parameterization, and also add insight to the in-sample versus out-of-sample question as it relates to crop insurance rating and distribution selection.

Figure 1: Comparison of Corn Yields at Different Aggregations over Time in Illinois

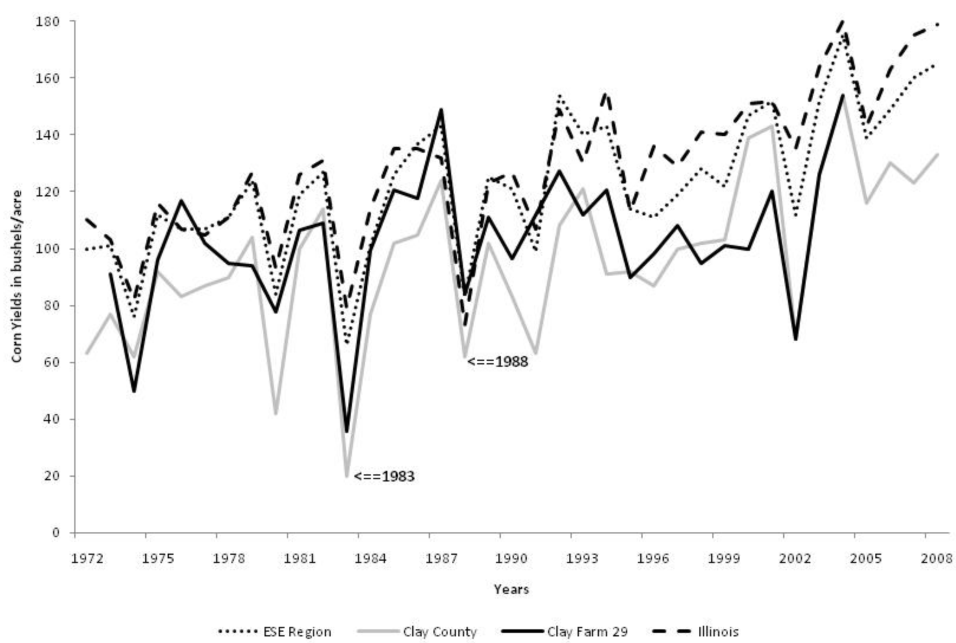
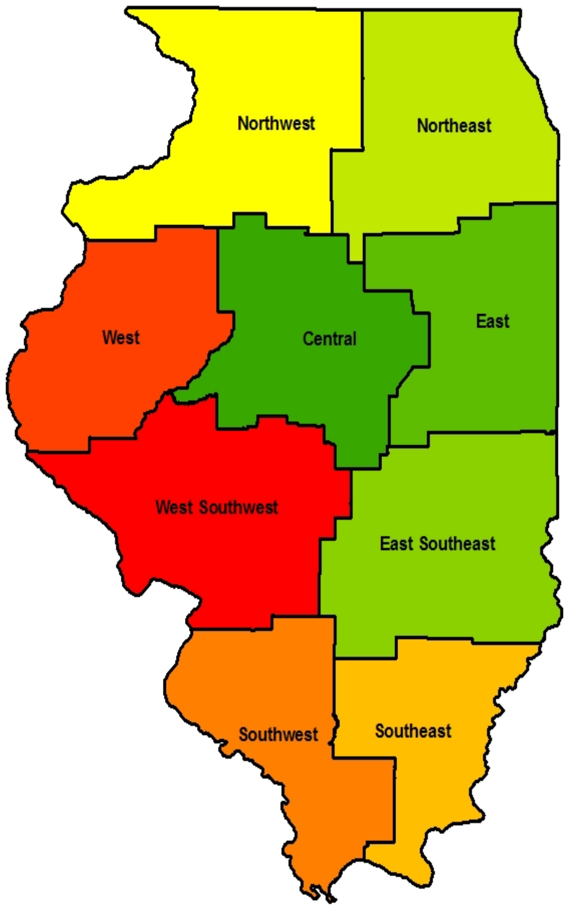


Figure 2: Map of Illinois NASS Crop Reporting Districts



Source: U.S. Census Bureau

Figure 3: Comparison of Distributional Forms on Representative Detrended FBFM Corn Farms

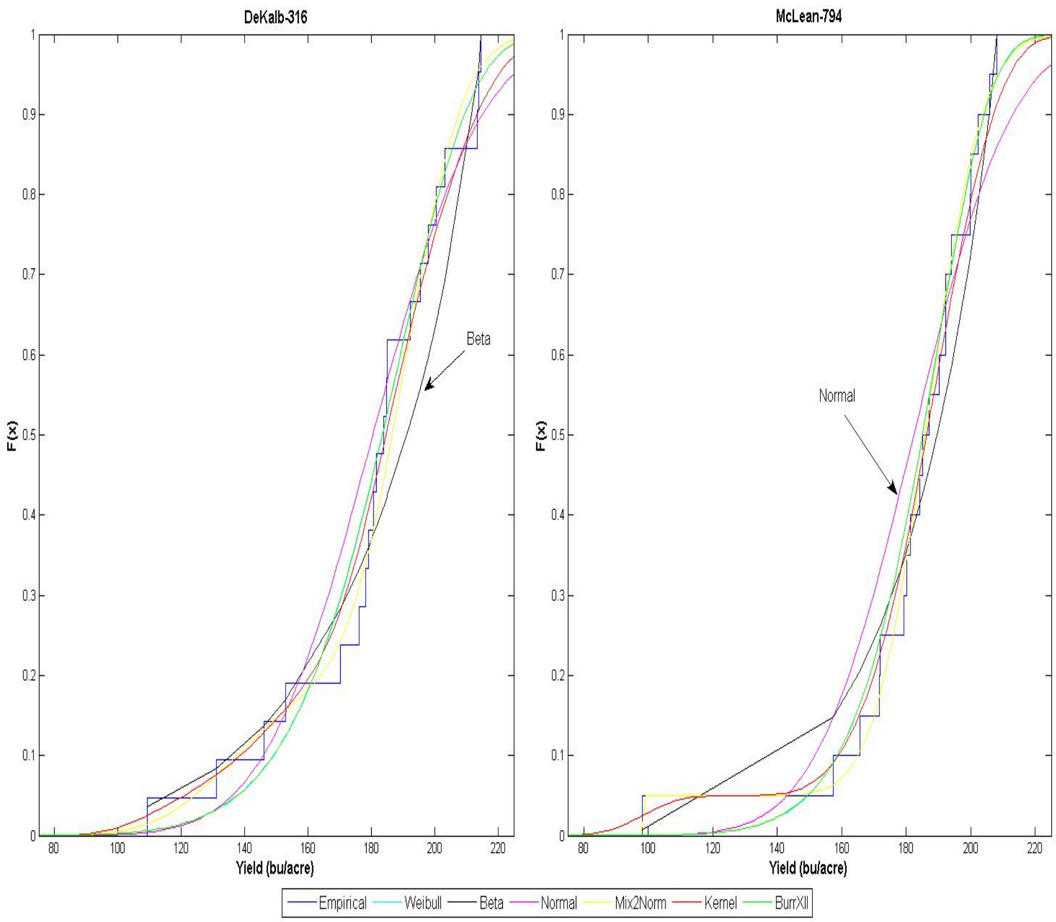


Figure 3: Comparison of Distributional Forms on Representative Detrended FBFM Corn Farms - Continued

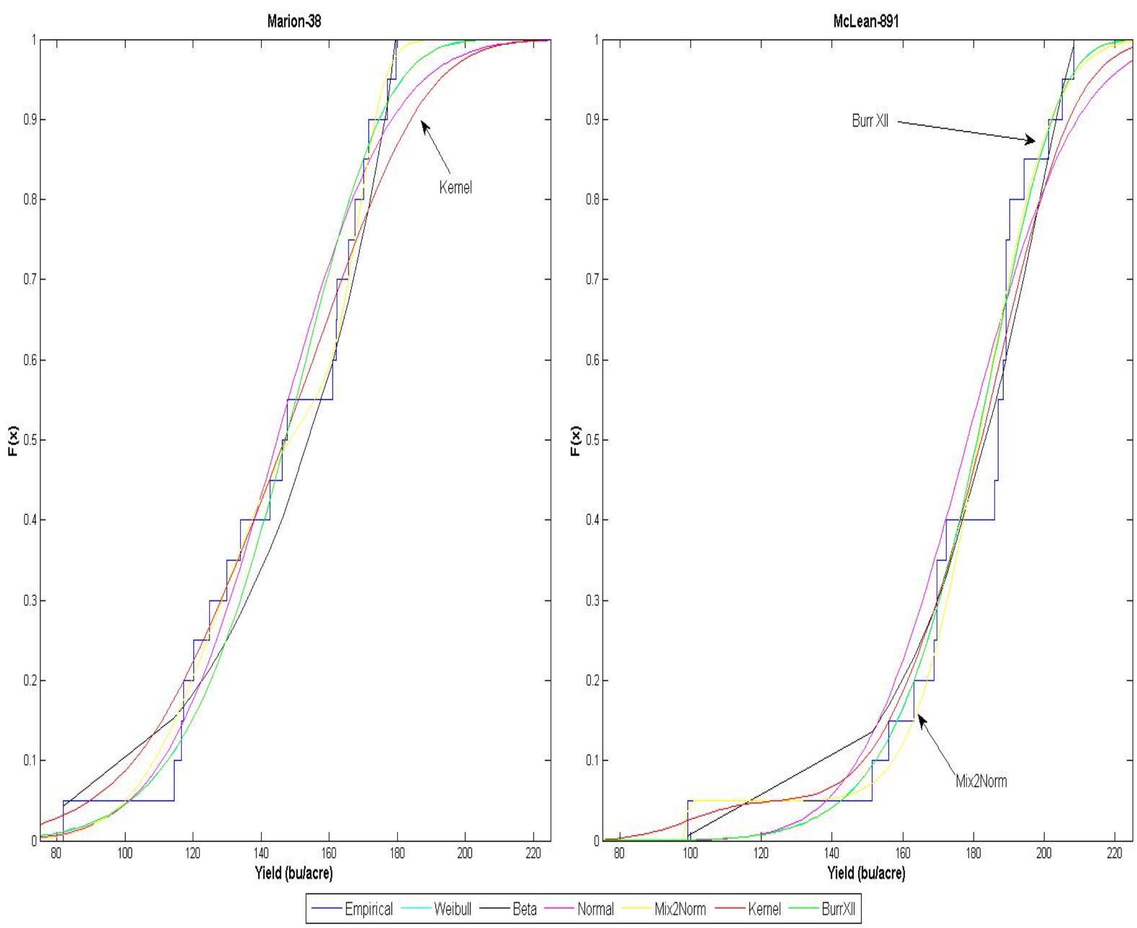


Figure 4: Comparison of Distributional Forms on Representative Detrended FBFM Soybean Farms

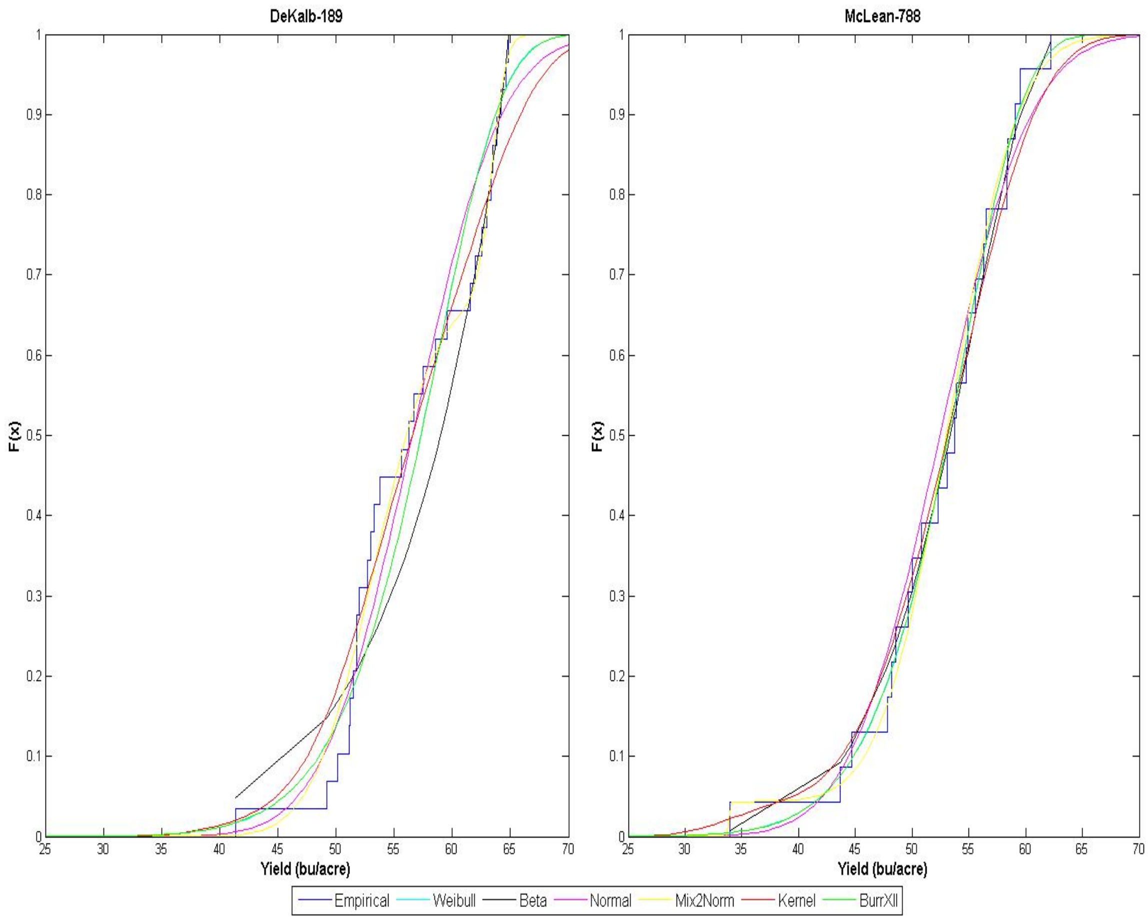


Figure 4: Comparison of Distributional Forms on Representative Detrended FBFM Soybean Farms - Continued

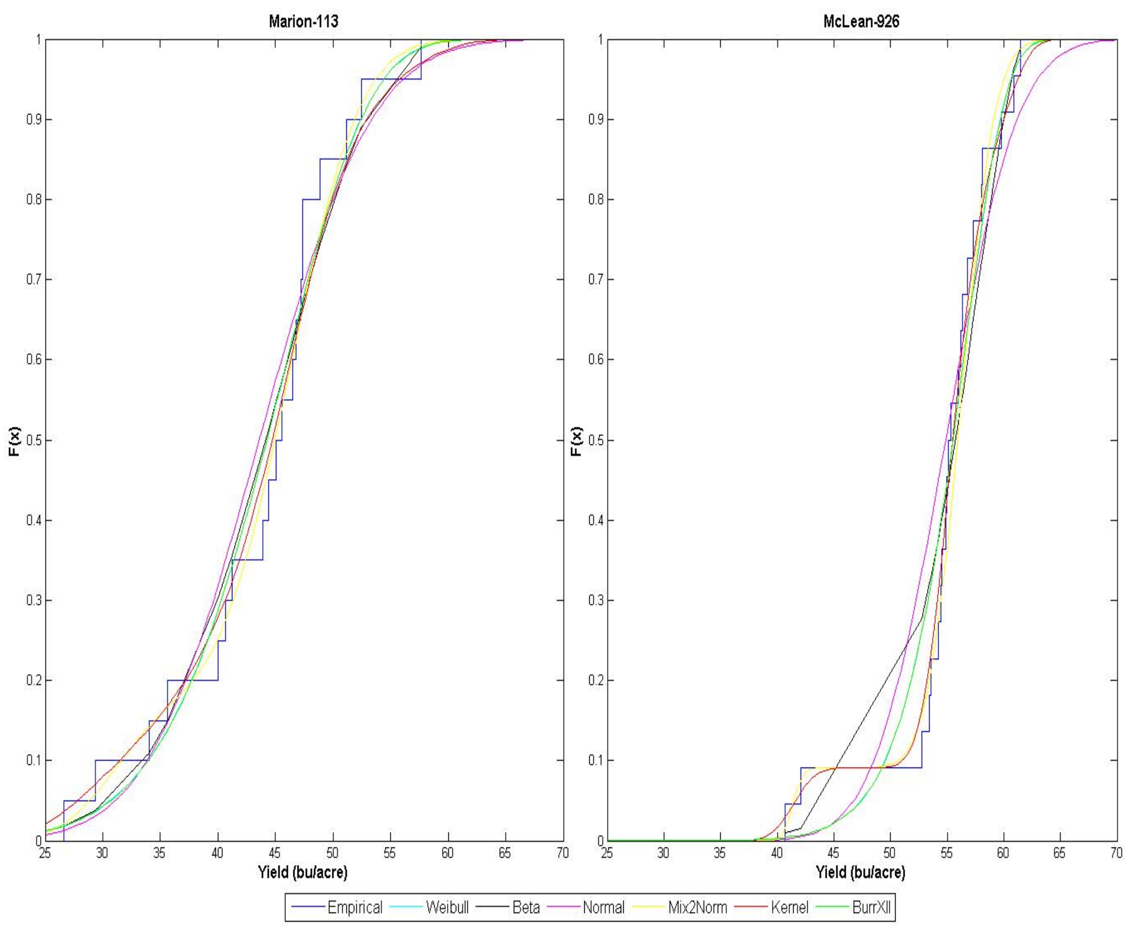


Table 1: FBFM Data: Number of Corn Farms with Twenty or More Years of Data with Sample Periods Ending in 2008

	Start of Sample Period**																		
Years of Data*	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**	**1978**	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**	**1985**	**1986**	**1987**	**1988**	**1989**	**Total Farms**
20+	719	244	134	120	74	74	88	85	69	53	57	56	39	68	75	42	60	31	2,088
21+	682	226	129	115	70	69	87	81	67	48	55	54	35	61	56	30	30		1,895
22+	653	217	124	112	69	69	85	78	63	47	50	53	29	42	39	20			1,750
23+	629	211	119	104	64	67	79	73	59	42	47	36	18	32	18				1,598
24+	607	202	116	101	59	64	74	68	52	39	44	26	13	10					1,475
25+	581	195	112	99	57	58	70	63	46	33	29	20	5						1,368
26+	550	188	105	92	53	52	65	52	37	23	17	9							1,243
27+	531	177	97	91	48	47	55	47	28	12	8								1,141
28+	495	169	87	85	42	39	43	31	19	3									1,013
29+	462	152	80	69	35	30	35	19	6										888
30+	428	132	75	56	28	20	24	5											768
31+	391	117	61	41	21	7	9												647
32+	334	99	46	25	12	2													518
33+	267	70	20	11	2														370
34+	197	47	6	4															254
35+	115	25	1																141
36+	50	5																	55
37+	16																		16

* Years of data in a sample period. Data may be continuous or non-continuous.

** Starting year of sample period to examine. Ending year of sample period is 2008.

Table 2: FBFM Data: Number of Soybean Farms with Twenty or More Years of Data with Sample Periods Ending in 2008

	Start of Sample Period[**]																		
Years of Data[*]	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**	**1978**	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**	**1985**	**1986**	**1987**	**1988**	**1989**	**Total Farms**
20+	560	239	129	112	61	79	84	81	60	50	58	56	38	68	79	42	55	30	1,881
21+	528	220	125	109	58	75	83	79	58	46	53	53	33	59	58	29	29		1,695
22+	512	210	120	104	56	73	80	77	55	45	50	49	27	42	40	18			1,558
23+	494	203	114	99	52	71	77	73	49	41	46	32	18	30	18				1,417
24+	478	195	112	93	49	69	70	66	45	37	42	20	13	12					1,301
25+	463	190	108	91	47	65	64	62	37	31	29	14	6						1,207
26+	437	182	102	84	45	59	61	51	32	23	16	5							1,097
27+	424	172	94	82	41	52	54	44	23	11	7								1,004
28+	394	162	87	76	35	42	43	31	15	3									888
29+	364	150	78	61	29	33	35	16	5										771
30+	332	130	71	52	24	20	22	6											657
31+	303	106	57	40	17	10	9												542
32+	260	89	42	24	9	3													427
33+	212	66	21	10	1														310
34+	158	46	4	4															212
35+	92	19	1																112
36+	37	5																	42
37+	13																		13

[*] Years of data in a sample period. Data may be continuous or non-continuous.

[**] Starting year of sample period to examine. Ending year of sample period is 2008.

Table 3: Illinois District Sample Characteristics from Filtered FBFM Corn Farms

		Sample Summary Statistics									
NASS District	**Data**	**Mean***	**St Dev***	**CV***	**Skew***	**Kurt***	**Max***	**Min***	**Farm Count**	**Yield Count**	**Avg % of Tot Acreage**
NW	**Original**	135.97	29.00	0.21	-0.42	3.72	242.00	15.50	395	10,959	19.50%
	Detrended - β= 2.07	173.81	23.97	0.14	-0.83	4.35	252.34	44.27			
NE	**Original**	141.39	28.68	0.20	-0.20	3.30	250.00	24.34	176	4,662	8.30%
	Detrended - β= 1.86	172.69	24.46	0.14	-0.51	3.36	265.10	46.03			
West	**Original**	136.35	32.75	0.24	-0.48	3.63	265.00	17.00	94	2,647	5.06%
	Detrended - β= 2.12	175.58	28.17	0.16	-0.60	3.60	275.27	48.37			
Central	**Original**	146.71	30.95	0.21	-0.63	3.80	263.00	17.35	519	14,292	25.75%
	Detrended - β= 2.03	181.82	26.23	0.14	-0.75	3.93	281.36	59.10			
East	**Original**	139.86	32.17	0.23	-0.75	3.64	239.00	18.96	296	7,849	13.90%
	Detrended - β= 1.84	171.49	28.37	0.17	-0.86	3.81	253.68	57.58			
WSW	**Original**	141.07	28.06	0.20	-0.39	3.33	238.00	27.02	151	4,106	6.93%
	Detrended - β= 1.87	177.59	24.45	0.14	-0.58	3.43	257.84	62.08			
ESE	**Original**	126.85	28.20	0.22	-0.35	3.27	218.00	15.00	253	6,682	12.17%
	Detrended - β= 1.61	155.54	26.73	0.17	-0.37	2.91	239.09	44.24			
SW	**Original**	106.05	28.80	0.27	-0.08	3.05	220.00	16.66	119	3,426	3.41%
	Detrended - β= 1.71	136.81	26.02	0.19	-0.30	2.89	228.57	29.00			
SE	**Original**	112.05	26.62	0.24	-0.18	3.20	206.64	16.84	85	2,277	4.97%
	Detrended - β= 1.68	142.55	24.77	0.17	-0.33	2.99	235.20	37.76			
Total	**Original**	136.25	29.80	0.22	-0.46	3.55	265.00	15.00	2,088	56,900	100%
	Detrended - β= 1.87	169.55	25.95	0.20	-0.66	3.69	277.79	30.10			

* in bushels/acre; for years 1972 to 2008

Table 4: Illinois District Sample Characteristics from Filtered FBFM Soybean Farms

		Sample Summary Statistics									
NASS District	**Data**	**Mean***	**St Dev***	**CV***	**Skew***	**Kurt***	**Max***	**Min***	**Farm Count**	**Yield Count**	**Avg % of Tot Acreage**
NW	**Original**	46.39	8.21	0.18	-0.56	3.81	89.00	7.00	240	6,442	8.88%
	Detrended - β= 0.42	53.61	7.65	0.14	-0.67	4.37	100.44	13.42			
NE	**Original**	45.07	7.27	0.16	-0.54	3.79	74.00	8.67	166	4,360	7.84%
	Detrended - β= 0.39	51.67	6.94	0.13	-0.65	4.14	81.24	16.76			
West	**Original**	44.21	7.30	0.17	-0.30	3.42	90.66	11.00	92	2,580	5.36%
	Detrended - β= 0.47	52.75	6.73	0.13	-0.32	3.29	100.91	17.40			
Central	**Original**	46.66	7.33	0.16	-0.85	4.45	73.00	7.27	503	13,783	26.98%
	Detrended - β= 0.44	54.30	6.86	0.13	-0.83	4.56	85.10	16.59			
East	**Original**	44.58	8.19	0.18	-0.74	3.91	89.00	9.00	293	7,772	16.49%
	Detrended - β= 0.44	52.19	7.65	0.15	-0.81	4.20	89.89	18.32			
WSW	**Original**	43.96	7.22	0.16	-0.57	3.65	75.00	10.92	150	4,071	8.02%
	Detrended - β= 0.40	51.68	6.56	0.13	-0.50	3.66	78.96	19.79			
ESE	**Original**	40.48	7.59	0.19	-0.44	3.58	76.00	7.00	249	6,568	15.61%
	Detrended - β= 0.43	48.23	7.05	0.15	-0.39	3.29	83.87	13.34			
SW	**Original**	37.52	8.55	0.23	-0.12	3.32	78.65	8.00	110	3,198	5.28%
	Detrended - β= 0.30	42.85	7.91	0.18	-0.18	3.31	86.66	8.59			
SE	**Original**	37.93	7.72	0.20	-0.18	3.11	73.00	7.87	78	2,099	5.53%
	Detrended - β= 0.40	45.05	7.23	0.16	-0.21	3.12	86.57	10.80			
Total	**Original**	44.11	7.68	0.17	-0.59	3.88	90.66	7.00	1,881	50,873	100%
	Detrended - β= 0.42	51.39	7.15	0.14	-0.63	4.02	100.28	8.83			

* in bushels/acre; for years 1972 to 2008

Table 5: Sample Average and Standard Deviation Sensitivity to Detrending Levels – FBFM Corn Farms

		Detrending Levels			
Statistics	**Original**	**State**	**District**	**County**	**Individual**
Average*	136.250	169.548	170.138	170.398	164.888
Standard Deviation*	29.802	25.951	25.904	25.900	24.954

* in bushels/acre.

Table 6: Sample Average and Standard Deviation Sensitivity to Detrending Levels – FBFM Soybean Farms

		Detrending Levels			
Statistics	**Original**	**State**	**District**	**County**	**Individual**
Average*	44.110	51.392	51.507	51.414	49.843
Standard Deviation*	7.682	7.146	7.162	7.157	6.900

* in bushels/acre.

Table 7: Goodness-of-Fit Results: Illinois Districts for Corn Farms

		Illinois Districts																													
		NW						NE						West						Central						East					
		Goodness-of-Fit Rankings*																													
Test	**Distribution**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
A-D[a]	**Weibull**	13%	49%	25%	7%	7%	0%	12%	40%	31%	8%	10%	0%	14%	39%	30%	7%	11%	0%	13%	48%	26%	6%	7%	0%	12%	51%	31%	3%	3%	0%
	gamma	1%	4%	2%	14%	79%	0%	2%	3%	1%	16%	77%	0%	2%	5%	2%	10%	81%	0%	1%	2%	1%	15%	80%	0%	0%	2%	1%	16%	81%	0%
	normal	9%	4%	15%	72%	0%	0%	8%	10%	15%	67%	0%	0%	9%	6%	16%	69%	0%	0%	8%	4%	17%	71%	0%	0%	6%	2%	18%	75%	0%	0%
	Burr XII	45%	37%	14%	3%	1%	0%	38%	39%	17%	6%	0%	0%	39%	39%	15%	8%	0%	0%	48%	36%	13%	3%	0%	0%	36%	42%	19%	2%	0%	0%
	invGauss	0%	0%	0%	0%	8%	92%	0%	0%	0%	0%	10%	90%	0%	0%	0%	0%	6%	94%	0%	0%	0%	0%	8%	92%	0%	0%	0%	0%	12%	88%
	beta	33%	6%	44%	4%	6%	8%	40%	8%	36%	3%	3%	10%	36%	12%	38%	6%	2%	6%	30%	9%	43%	5%	4%	8%	46%	3%	31%	4%	3%	12%
K-S[b]	**Weibull**	16%	39%	31%	7%	7%	0%	18%	39%	26%	7%	10%	0%	18%	32%	34%	2%	14%	0%	18%	39%	26%	7%	9%	0%	16%	45%	30%	6%	3%	0%
	gamma	6%	4%	1%	9%	80%	0%	2%	8%	3%	10%	77%	0%	6%	9%	3%	3%	78%	0%	4%	6%	2%	9%	79%	0%	2%	1%	1%	7%	89%	0%
	normal	8%	8%	13%	71%	0%	0%	15%	6%	15%	64%	0%	0%	12%	6%	7%	75%	0%	0%	8%	8%	15%	68%	0%	0%	5%	6%	10%	79%	0%	0%
	Burr XII	34%	42%	16%	6%	1%	0%	27%	40%	21%	9%	2%	0%	30%	43%	15%	11%	1%	0%	39%	40%	14%	6%	2%	0%	34%	43%	21%	2%	0%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	37%	5%	39%	7%	12%	0%	38%	6%	35%	9%	12%	0%	34%	10%	41%	8%	7%	0%	31%	7%	43%	10%	10%	0%	43%	5%	39%	6%	8%	0%
χ^{2c}	**Weibull**	17%	31%	28%	14%	9%	0%	20%	31%	29%	9%	11%	0%	18%	25%	20%	19%	17%	0%	17%	29%	30%	12%	12%	0%	12%	31%	33%	14%	10%	0%
	gamma	12%	11%	9%	13%	55%	0%	12%	10%	5%	14%	60%	0%	19%	8%	12%	7%	53%	0%	10%	10%	7%	12%	60%	0%	9%	9%	8%	11%	63%	0%
	normal	14%	15%	13%	47%	11%	0%	9%	14%	16%	51%	10%	0%	11%	27%	10%	40%	11%	0%	13%	15%	14%	49%	9%	0%	12%	15%	12%	52%	9%	0%
	Burr XII	23%	34%	20%	16%	7%	0%	21%	40%	21%	15%	3%	0%	16%	26%	28%	20%	9%	0%	19%	38%	22%	15%	6%	0%	21%	40%	21%	14%	5%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	34%	9%	29%	10%	18%	0%	38%	6%	30%	10%	16%	0%	35%	14%	28%	14%	9%	0%	40%	8%	28%	12%	12%	0%	46%	4%	26%	10%	13%	0%
Weighted[d]	**Weibull**	18%	40%	27%	10%	5%	0%	18%	38%	26%	9%	8%	0%	22%	25%	33%	7%	14%	0%	17%	42%	27%	7%	7%	0%	16%	42%	31%	8%	2%	0%
	gamma	3%	3%	4%	15%	74%	0%	2%	3%	5%	19%	71%	0%	7%	5%	3%	18%	67%	0%	2%	4%	3%	15%	77%	0%	1%	3%	0%	12%	84%	0%
	normal	9%	8%	19%	58%	7%	0%	10%	7%	24%	53%	5%	0%	8%	12%	16%	55%	9%	0%	8%	7%	21%	59%	5%	0%	6%	4%	21%	64%	5%	0%
	Burr XII	46%	32%	14%	7%	1%	0%	38%	38%	12%	11%	1%	0%	40%	32%	14%	12%	2%	0%	48%	32%	12%	7%	1%	0%	42%	38%	16%	3%	0%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	24%	17%	35%	10%	13%	0%	31%	13%	34%	8%	14%	0%	24%	26%	34%	8%	8%	0%	25%	16%	38%	12%	10%	0%	35%	13%	31%	12%	9%	0%

Continued on next page

Table 7 – continued from previous page

		Illinois Districts																													
		WSW						ESE						SW						SE						Total					
		Goodness-of-Fit Rankings																													
Test	**Distribution**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
A-D[a]	**Weibull**	11%	47%	25%	7%	10%	0%	8%	41%	22%	17%	12%	0%	13%	35%	27%	13%	13%	0%	5%	41%	29%	11%	14%	0%	12%	45%	27%	8%	8%	0%
	gamma	3%	6%	1%	21%	69%	0%	4%	8%	5%	18%	65%	0%	5%	5%	5%	12%	75%	0%	5%	3%	1%	16%	75%	0%	2%	4%	2%	16%	77%	0%
	normal	14%	10%	19%	56%	0%	0%	20%	13%	19%	48%	0%	0%	18%	14%	14%	55%	0%	0%	14%	12%	13%	61%	0%	0%	11%	7%	16%	66%	0%	0%
	Burr XII	40%	31%	21%	6%	1%	0%	37%	29%	24%	8%	2%	0%	34%	37%	20%	9%	0%	0%	28%	34%	32%	7%	0%	0%	41%	36%	17%	5%	1%	0%
	invGauss	0%	0%	0%	0%	13%	87%	0%	0%	0%	0%	11%	89%	0%	0%	0%	0%	5%	95%	0%	0%	0%	0%	5%	95%	0%	0%	0%	0%	9%	91%
	beta	32%	5%	34%	9%	7%	13%	31%	10%	30%	9%	9%	11%	31%	10%	35%	12%	7%	5%	47%	11%	25%	7%	5%	5%	35%	8%	37%	6%	5%	9%
K-S[b]	**Weibull**	15%	40%	24%	9%	13%	0%	17%	34%	26%	12%	11%	0%	23%	25%	23%	16%	12%	0%	11%	25%	33%	12%	20%	0%	17%	38%	28%	8%	9%	0%
	gamma	7%	9%	3%	14%	68%	0%	8%	11%	3%	14%	64%	0%	11%	6%	4%	6%	73%	0%	7%	12%	7%	11%	64%	0%	5%	6%	2%	9%	77%	0%
	normal	14%	10%	16%	59%	0%	0%	17%	13%	22%	48%	0%	0%	20%	10%	17%	53%	0%	0%	21%	14%	7%	58%	0%	0%	11%	9%	14%	66%	0%	0%
	Burr XII	31%	36%	19%	12%	2%	0%	26%	36%	23%	13%	2%	0%	21%	43%	19%	14%	3%	0%	21%	39%	21%	17%	1%	0%	32%	40%	18%	8%	2%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	33%	6%	38%	6%	17%	0%	33%	6%	25%	13%	22%	0%	25%	15%	37%	10%	13%	0%	41%	9%	33%	3%	14%	0%	35%	7%	38%	8%	12%	0%
χ^2[c]	**Weibull**	13%	28%	27%	16%	16%	0%	22%	25%	28%	13%	13%	0%	23%	28%	19%	14%	16%	0%	11%	25%	38%	18%	8%	0%	17%	29%	29%	14%	12%	0%
	gamma	17%	17%	6%	14%	45%	0%	17%	10%	7%	13%	53%	0%	14%	16%	7%	14%	49%	0%	13%	14%	5%	18%	49%	0%	13%	11%	7%	13%	56%	0%
	normal	17%	19%	14%	37%	13%	0%	11%	18%	21%	42%	8%	0%	19%	14%	16%	42%	9%	0%	22%	12%	4%	47%	14%	0%	13%	16%	14%	47%	10%	0%
	Burr XII	20%	31%	20%	20%	9%	0%	19%	36%	19%	22%	3%	0%	22%	32%	16%	21%	9%	0%	13%	38%	24%	13%	12%	0%	20%	36%	21%	17%	6%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	33%	4%	33%	12%	18%	0%	31%	11%	26%	9%	23%	0%	22%	10%	41%	10%	17%	0%	41%	11%	29%	3%	17%	0%	37%	8%	29%	10%	16%	0%
Weighted[d]	**Weibull**	19%	39%	22%	10%	11%	0%	17%	36%	20%	17%	10%	0%	22%	29%	24%	13%	13%	0%	12%	28%	29%	16%	16%	0%	18%	38%	27%	10%	8%	0%
	gamma	4%	6%	4%	21%	66%	0%	6%	9%	6%	16%	63%	0%	5%	6%	9%	14%	65%	0%	5%	8%	5%	14%	67%	0%	3%	5%	4%	15%	73%	0%
	normal	15%	9%	23%	48%	5%	0%	19%	10%	25%	42%	4%	0%	22%	13%	13%	50%	4%	0%	22%	13%	17%	43%	4%	0%	11%	8%	21%	55%	5%	0%
	Burr XII	40%	28%	19%	12%	1%	0%	31%	31%	25%	11%	2%	0%	28%	39%	14%	14%	5%	0%	17%	39%	26%	16%	1%	0%	41%	34%	16%	8%	1%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	22%	19%	32%	9%	18%	0%	28%	14%	24%	14%	21%	0%	23%	14%	40%	10%	14%	0%	43%	12%	22%	11%	12%	0%	27%	16%	33%	11%	13%	0%

[*] Highlighted values represent the distributional form that fits the highest percentage of farms by ranking and goodness-of-fit test.
[a] Anderson-Darling Test - (Stephens, 1974)
[b] Kolmogorov-Smirnov Test - (Chakravart et al., 1967)
[c] χ^2 Test - (Snedecor and Cochran, 1989)
[d] Weighted Test = (.334*A-D + .333*K-S +.333*χ^2)

Table 8: Goodness-of-Fit Results: Illinois Districts for Soybean Farms

		Illinois Districts																													
		NW						NE						West						Central						East					
		Goodness-of-Fit Rankings*																													
Test	**Distribution**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
A-D[a]	**Weibull**	13%	51%	18%	7%	11%	0%	15%	44%	17%	11%	12%	0%	13%	39%	19%	9%	20%	0%	14%	54%	16%	7%	9%	0%	14%	52%	21%	7%	6%	0%
	gamma	0%	6%	2%	15%	76%	0%	2%	8%	1%	14%	75%	0%	3%	11%	1%	19%	66%	0%	1%	3%	1%	12%	82%	0%	2%	1%	1%	20%	75%	0%
	normal	11%	8%	18%	64%	0%	0%	14%	6%	16%	64%	0%	0%	18%	8%	20%	54%	0%	0%	6%	7%	14%	74%	0%	0%	4%	6%	22%	67%	0%	0%
	Burr XII	50%	30%	14%	5%	0%	0%	49%	32%	14%	5%	0%	0%	44%	32%	19%	5%	0%	0%	59%	27%	12%	2%	0%	0%	53%	33%	11%	2%	1%	0%
	invGauss	0%	0%	0%	0%	9%	91%	0%	0%	0%	0%	10%	90%	0%	0%	0%	0%	9%	91%	0%	0%	0%	0%	7%	93%	0%	0%	0%	0%	16%	84%
	beta	26%	6%	48%	8%	4%	9%	20%	9%	52%	6%	3%	10%	23%	10%	41%	13%	5%	9%	21%	10%	57%	4%	2%	7%	27%	8%	44%	4%	2%	16%
K-S[b]	**Weibull**	20%	38%	23%	8%	11%	0%	14%	40%	19%	15%	13%	0%	18%	28%	24%	15%	15%	0%	18%	45%	19%	7%	11%	0%	17%	44%	25%	7%	6%	0%
	gamma	5%	6%	1%	8%	80%	0%	4%	13%	3%	6%	73%	0%	6%	15%	5%	5%	68%	0%	4%	6%	2%	8%	80%	0%	3%	3%	1%	8%	85%	0%
	normal	8%	11%	13%	68%	0%	0%	15%	9%	13%	63%	0%	0%	19%	9%	18%	54%	0%	0%	8%	7%	12%	73%	0%	0%	4%	6%	16%	73%	0%	0%
	Burr XII	42%	39%	10%	7%	2%	0%	40%	28%	22%	8%	2%	0%	34%	33%	15%	11%	6%	0%	48%	32%	13%	5%	1%	0%	42%	37%	16%	4%	1%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	26%	7%	53%	8%	6%	0%	27%	10%	42%	9%	12%	0%	23%	15%	38%	14%	10%	0%	23%	9%	54%	7%	8%	0%	34%	10%	41%	7%	8%	0%
χ^2[c]	**Weibull**	19%	32%	22%	17%	10%	0%	21%	31%	27%	10%	11%	0%	25%	28%	23%	10%	14%	0%	21%	32%	24%	12%	11%	0%	19%	33%	27%	11%	9%	0%
	gamma	12%	11%	6%	15%	55%	0%	15%	12%	5%	14%	54%	0%	20%	3%	8%	13%	57%	0%	8%	11%	7%	11%	63%	0%	7%	10%	5%	14%	65%	0%
	normal	13%	14%	17%	43%	13%	0%	13%	15%	13%	50%	9%	0%	9%	16%	14%	51%	10%	0%	11%	14%	13%	52%	10%	0%	13%	12%	15%	51%	10%	0%
	Burr XII	25%	30%	24%	13%	7%	0%	25%	33%	18%	15%	8%	0%	22%	43%	20%	13%	3%	0%	31%	34%	17%	14%	4%	0%	25%	35%	21%	12%	6%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	30%	12%	31%	11%	15%	0%	25%	9%	38%	10%	17%	0%	24%	10%	35%	14%	16%	0%	29%	9%	39%	11%	12%	0%	36%	10%	32%	12%	10%	0%
Weighted[d]	**Weibull**	19%	38%	23%	10%	10%	0%	19%	39%	20%	10%	12%	0%	22%	33%	18%	11%	16%	0%	19%	45%	22%	7%	8%	0%	20%	43%	23%	9%	5%	0%
	gamma	3%	5%	2%	19%	70%	0%	6%	8%	5%	10%	72%	0%	5%	10%	6%	13%	66%	0%	2%	4%	3%	13%	78%	0%	3%	1%	1%	14%	80%	0%
	normal	11%	8%	18%	57%	6%	0%	15%	10%	16%	56%	2%	0%	16%	9%	23%	49%	3%	0%	8%	6%	19%	62%	5%	0%	5%	8%	20%	64%	4%	0%
	Burr XII	46%	35%	10%	6%	3%	0%	44%	29%	17%	8%	1%	0%	39%	32%	18%	10%	1%	0%	55%	29%	9%	6%	1%	0%	51%	29%	14%	5%	0%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	21%	13%	47%	9%	10%	0%	16%	13%	43%	15%	13%	0%	18%	16%	35%	16%	14%	0%	17%	16%	47%	12%	8%	0%	21%	19%	42%	8%	10%	0%

Continued on next page

Table 8 – continued from previous page

		Illinois Districts																													
		WSW						ESE						SW						SE						Total					
		Goodness-of-Fit Rankings																													
Test	**Distribution**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
A-D[a]	**Weibull**	14%	36%	25%	7%	18%	0%	10%	39%	28%	11%	12%	0%	9%	35%	22%	19%	15%	0%	9%	36%	27%	14%	14%	0%	13%	47%	20%	9%	11%	0%
	gamma	0%	4%	1%	20%	74%	0%	4%	2%	2%	18%	73%	0%	7%	1%	3%	14%	74%	0%	0%	8%	6%	18%	68%	0%	2%	4%	2%	16%	76%	0%
	normal	12%	10%	14%	64%	0%	0%	12%	10%	20%	58%	0%	0%	11%	23%	17%	49%	0%	0%	23%	8%	17%	53%	0%	0%	10%	8%	17%	65%	0%	0%
	Burr XII	50%	34%	11%	5%	0%	0%	42%	37%	14%	6%	0%	0%	40%	31%	22%	5%	1%	0%	35%	33%	24%	8%	0%	0%	50%	31%	14%	4%	0%	0%
	invGauss	0%	0%	0%	0%	5%	95%	0%	0%	0%	0%	8%	92%	0%	0%	0%	0%	1%	99%	0%	0%	0%	0%	9%	91%	0%	0%	0%	0%	9%	91%
	beta	23%	16%	49%	4%	3%	5%	31%	12%	36%	7%	6%	8%	33%	10%	35%	13%	9%	1%	33%	15%	26%	8%	9%	9%	25%	10%	47%	6%	4%	9%
K-S[b]	**Weibull**	9%	40%	25%	12%	14%	0%	19%	34%	26%	10%	11%	0%	16%	36%	17%	14%	17%	0%	14%	24%	27%	21%	14%	0%	17%	40%	22%	10%	11%	0%
	gamma	5%	7%	3%	7%	78%	0%	8%	5%	2%	12%	73%	0%	9%	9%	6%	10%	67%	0%	8%	15%	3%	12%	62%	0%	5%	7%	3%	8%	77%	0%
	normal	9%	9%	14%	68%	0%	0%	11%	16%	14%	60%	0%	0%	16%	16%	19%	49%	0%	0%	15%	14%	18%	53%	0%	0%	10%	10%	14%	67%	0%	0%
	Burr XII	43%	31%	15%	9%	1%	0%	29%	37%	23%	8%	4%	0%	32%	32%	21%	10%	5%	0%	36%	32%	15%	8%	9%	0%	41%	34%	16%	7%	2%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	34%	12%	43%	4%	7%	0%	34%	8%	35%	11%	12%	0%	28%	7%	36%	18%	11%	0%	27%	15%	36%	6%	15%	0%	28%	10%	45%	8%	9%	0%
χ^{2c}	**Weibull**	20%	23%	24%	18%	15%	0%	24%	25%	21%	13%	16%	0%	21%	18%	21%	19%	20%	0%	18%	26%	23%	14%	20%	0%	21%	29%	24%	13%	12%	0%
	gamma	18%	12%	7%	8%	55%	0%	17%	11%	6%	13%	53%	0%	22%	17%	6%	3%	51%	0%	12%	8%	11%	11%	59%	0%	12%	11%	6%	12%	58%	0%
	normal	9%	16%	18%	47%	9%	0%	12%	21%	12%	47%	8%	0%	13%	21%	23%	34%	9%	0%	21%	15%	15%	44%	5%	0%	12%	15%	15%	48%	10%	0%
	Burr XII	21%	35%	16%	17%	11%	0%	17%	34%	23%	20%	6%	0%	17%	28%	23%	24%	7%	0%	17%	38%	27%	15%	3%	0%	24%	34%	20%	15%	6%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	31%	15%	35%	9%	9%	0%	30%	8%	38%	7%	17%	0%	27%	16%	26%	19%	13%	0%	32%	14%	24%	17%	14%	0%	30%	11%	35%	11%	13%	0%
Weighted[d]	**Weibull**	14%	34%	26%	11%	15%	0%	19%	32%	27%	15%	7%	0%	13%	29%	28%	17%	14%	0%	14%	30%	24%	21%	11%	0%	18%	39%	23%	11%	9%	0%
	gamma	3%	7%	3%	16%	72%	0%	5%	4%	4%	16%	71%	0%	6%	6%	7%	18%	62%	0%	2%	8%	12%	12%	67%	0%	3%	5%	4%	14%	74%	0%
	normal	12%	8%	20%	54%	6%	0%	15%	12%	21%	46%	5%	0%	16%	16%	24%	38%	5%	0%	18%	17%	24%	39%	2%	0%	11%	9%	20%	56%	5%	0%
	Burr XII	48%	31%	7%	10%	3%	0%	34%	37%	16%	10%	3%	0%	34%	33%	16%	13%	4%	0%	41%	27%	14%	14%	5%	0%	47%	31%	12%	8%	2%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	23%	20%	44%	9%	5%	0%	27%	15%	31%	14%	14%	0%	31%	16%	24%	14%	15%	0%	26%	18%	26%	14%	17%	0%	21%	16%	41%	12%	11%	0%

[*] Highlighted values represent the distributional form that fits the highest percentage of farms by ranking and goodness-of-fit test.
[a] Anderson-Darling Test - (Stephens, 1974)
[b] Kolmogorov-Smirnov Test - (Chakravart et al., 1967)
[c] χ^2 Test - (Snedecor and Cochran, 1989)
[d] Weighted Test = (.334*A-D + .333*K-S +.333*χ^2)

Table 9: Heat-Map of Combined Goodness-of-Fit Ranking Scores: Illinois Districts for Corn Farms

		Illinois Districts*									
		NW	**NE**	**West**	**Central**	**East**	**WSW**	**ESE**	**SW**	**SE**	**Total**
A-D	**Weibull**	2.454	2.656	2.636	2.474	2.336	2.586	2.839	2.775	2.882	2.557
	gamma	4.673	4.636	4.625	4.705	4.750	4.471	4.326	4.477	4.526	4.614
	normal	3.504	3.403	3.455	3.498	3.614	3.179	2.949	3.054	3.197	3.380
	Burr XII	1.776	1.903	1.920	1.724	1.886	1.964	2.097	2.036	2.171	1.877
	invGauss	5.921	5.903	5.943	5.919	5.875	5.871	5.886	5.946	5.947	5.908
	beta	2.673	2.500	2.420	2.681	2.539	2.929	2.903	2.712	2.276	2.664
K-S	**Weibull**	2.493	2.513	2.614	2.488	2.350	2.643	2.669	2.676	3.053	2.542
	gamma	4.525	4.513	4.398	4.549	4.804	4.264	4.169	4.243	4.145	4.472
	normal	3.472	3.286	3.443	3.439	3.614	3.207	3.017	3.036	3.013	3.352
	Burr XII	1.984	2.182	2.114	1.921	1.921	2.193	2.292	2.351	2.382	2.069
	invGauss	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
	beta	2.525	2.506	2.432	2.604	2.311	2.693	2.852	2.694	2.408	2.565
χ^2	**Weibull**	2.670	2.604	2.920	2.726	2.764	2.943	2.699	2.712	2.882	2.737
	gamma	3.881	4.006	3.670	4.018	4.111	3.529	3.750	3.658	3.750	3.890
	normal	3.261	3.403	3.125	3.246	3.311	3.093	3.191	3.090	3.197	3.237
	Burr XII	2.493	2.370	2.807	2.514	2.425	2.657	2.538	2.631	2.724	2.527
	invGauss	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
	beta	2.694	2.617	2.477	2.496	2.389	2.779	2.822	2.910	2.447	2.609
Weighted	**Weibull**	2.449	2.513	2.659	2.451	2.386	2.557	2.674	2.658	2.961	2.521
	gamma	4.522	4.545	4.341	4.614	4.754	4.379	4.216	4.270	4.303	4.502
	normal	3.462	3.357	3.443	3.465	3.579	3.186	3.025	3.009	2.934	3.352
	Burr XII	1.850	1.974	2.057	1.805	1.807	2.057	2.216	2.297	2.447	1.959
	invGauss	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
	beta	2.718	2.610	2.500	2.665	2.475	2.821	2.869	2.766	2.355	2.666

* Darker colors represent larger values and lighter colors represent smaller values.

Table 10: Heat-Map of Combined Goodness-of-Fit Ranking Scores: Illinois Districts for Soybean Farms

		Illinois Districts*									
		NW	**NE**	**West**	**Central**	**East**	**WSW**	**ESE**	**SW**	**SE**	**Total**
A-D	**Weibull**	2.522	2.601	2.848	2.443	2.398	2.803	2.746	2.968	2.864	2.591
	gamma	4.608	4.510	4.342	4.711	4.656	4.635	4.531	4.468	4.470	4.603
	normal	3.354	3.294	3.114	3.557	3.523	3.285	3.250	3.043	3.000	3.372
	Burr XII	1.761	1.748	1.848	1.580	1.638	1.708	1.862	1.957	2.045	1.724
	invGauss	5.914	5.902	5.911	5.932	5.842	5.949	5.920	5.989	5.909	5.914
	beta	2.842	2.944	2.937	2.776	2.943	2.620	2.692	2.574	2.712	2.796
K-S	**Weibull**	2.541	2.720	2.823	2.481	2.423	2.803	2.594	2.798	2.970	2.592
	gamma	4.517	4.308	4.139	4.549	4.688	4.453	4.375	4.181	4.061	4.459
	normal	3.435	3.245	3.076	3.496	3.599	3.409	3.228	3.011	3.091	3.380
	Burr XII	1.885	2.042	2.228	1.802	1.835	1.956	2.223	2.245	2.212	1.965
	invGauss	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
	beta	2.622	2.685	2.734	2.673	2.455	2.380	2.580	2.766	2.667	2.604
χ^2	**Weibull**	2.679	2.587	2.595	2.589	2.581	2.869	2.714	2.989	2.909	2.672
	gamma	3.890	3.790	3.835	4.103	4.197	3.708	3.741	3.436	3.970	3.933
	normal	3.282	3.273	3.367	3.359	3.330	3.299	3.170	3.043	2.955	3.275
	Burr XII	2.464	2.490	2.316	2.278	2.384	2.613	2.652	2.777	2.500	2.450
	invGauss	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
	beta	2.684	2.860	2.886	2.671	2.509	2.511	2.723	2.755	2.667	2.670
Weighted	**Weibull**	2.541	2.573	2.684	2.411	2.369	2.781	2.589	2.904	2.848	2.544
	gamma	4.474	4.350	4.241	4.622	4.670	4.467	4.424	4.223	4.348	4.500
	normal	3.388	3.189	3.127	3.496	3.552	3.328	3.143	3.011	2.894	3.339
	Burr XII	1.852	1.937	2.025	1.684	1.735	1.883	2.112	2.202	2.136	1.868
	invGauss	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6.000
	beta	2.746	2.951	2.924	2.787	2.674	2.540	2.732	2.660	2.773	2.749

* Darker colors represent larger values and lighter colors represent smaller values.

Table 11: Goodness-of-Fit Results Grouped by Sample Size: Corn Farms

		Bins																	
		$20 \leq SampleSize < 24$						$24 \leq SampleSize < 30$						$30 \leq SampleSize \leq 37$					
		Goodness-of-Fit Rankings*																	
Test	**Distribution**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
A-D[a]	**Weibull**	12%	50%	24%	7%	8%	0%	12%	43%	26%	10%	9%	0%	12%	43%	30%	6%	8%	0%
	gamma	3%	5%	1%	22%	70%	0%	1%	5%	3%	15%	77%	0%	2%	2%	2%	10%	83%	0%
	normal	11%	6%	22%	60%	0%	0%	11%	9%	16%	65%	0%	0%	10%	5%	11%	74%	0%	0%
	BurrXII	46%	33%	16%	5%	0%	0%	40%	36%	19%	5%	1%	0%	37%	41%	17%	4%	0%	0%
	invGauss	0%	0%	0%	0%	14%	86%	0%	0%	0%	0%	9%	91%	0%	0%	0%	0%	5%	95%
	beta	28%	6%	36%	6%	8%	14%	37%	8%	37%	6%	4%	9%	40%	9%	39%	5%	3%	5%
K-S[b]	**Weibull**	18%	37%	27%	9%	9%	0%	16%	38%	29%	7%	10%	0%	17%	39%	28%	8%	9%	0%
	gamma	5%	7%	3%	12%	73%	0%	5%	6%	2%	9%	77%	0%	5%	6%	2%	7%	81%	0%
	normal	12%	9%	17%	62%	0%	0%	11%	10%	14%	66%	0%	0%	10%	7%	12%	71%	0%	0%
	BurrXII	31%	42%	19%	7%	1%	0%	33%	39%	17%	10%	2%	0%	32%	40%	18%	8%	1%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	34%	6%	34%	10%	17%	0%	35%	6%	38%	9%	11%	0%	36%	8%	41%	7%	9%	0%
χ^{2}[c]	**Weibull**	22%	27%	31%	11%	9%	0%	15%	25%	30%	15%	15%	0%	15%	35%	24%	15%	11%	0%
	gamma	12%	10%	5%	11%	62%	0%	15%	11%	9%	13%	52%	0%	10%	13%	8%	14%	55%	0%
	normal	13%	14%	14%	54%	5%	0%	15%	19%	14%	41%	10%	0%	11%	15%	13%	45%	16%	0%
	BurrXII	19%	44%	20%	14%	3%	0%	16%	35%	22%	20%	7%	0%	26%	29%	21%	15%	8%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	34%	6%	30%	9%	21%	0%	39%	9%	24%	11%	17%	0%	37%	8%	33%	11%	10%	0%
Weighted[d]	**Weibull**	20%	40%	24%	9%	7%	0%	15%	37%	29%	10%	9%	0%	17%	38%	26%	11%	7%	0%
	gamma	4%	4%	3%	17%	72%	0%	2%	6%	6%	15%	71%	0%	3%	4%	3%	14%	75%	0%
	normal	10%	8%	27%	52%	2%	0%	13%	8%	20%	53%	5%	0%	9%	8%	15%	60%	8%	0%
	BurrXII	42%	37%	14%	7%	1%	0%	40%	31%	16%	11%	1%	0%	41%	33%	17%	7%	2%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	24%	12%	32%	15%	18%	0%	29%	18%	29%	10%	13%	0%	29%	17%	39%	7%	8%	0%

* Highlighted values represent the distributional form that fits the highest percentage of farms by ranking and goodness-of-fit test.

[a] Anderson-Darling Test - (Stephens, 1974)

[b] Kolmogorov-Smirnov Test - (Chakravart et al., 1967)

[c] χ^2 Test - (Snedecor and Cochran, 1989)

[d] Weighted Test = (.334*A-D + .333*K-S +.333*χ^2)

Table 12: Goodness-of-Fit Results Grouped by Sample Size: Soybean Farms

		Bins																	
		$20 \leq SampleSize < 24$						$24 \leq SampleSize < 30$						$30 \leq SampleSize \leq 37$					
		Goodness-of-Fit Rankings*																	
Test	**Distribution**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
A-D[a]	**Weibull**	12%	48%	20%	8%	12%	0%	14%	43%	20%	10%	12%	0%	12%	49%	21%	9%	9%	0%
	gamma	2%	6%	1%	23%	67%	0%	2%	3%	3%	14%	78%	0%	2%	2%	1%	10%	84%	0%
	normal	13%	7%	23%	56%	0%	0%	9%	9%	16%	67%	0%	0%	7%	10%	11%	73%	0%	0%
	BurrXII	53%	30%	12%	6%	0%	0%	47%	34%	15%	4%	0%	0%	52%	29%	16%	2%	0%	0%
	invGauss	0%	0%	0%	0%	15%	85%	0%	0%	0%	0%	7%	93%	0%	0%	0%	0%	4%	96%
	beta	20%	9%	44%	6%	6%	15%	28%	10%	46%	6%	3%	7%	27%	11%	51%	5%	3%	4%
K-S[b]	**Weibull**	18%	38%	23%	9%	13%	0%	15%	42%	21%	9%	13%	0%	17%	40%	23%	11%	8%	0%
	gamma	7%	7%	2%	11%	73%	0%	5%	8%	3%	8%	77%	0%	3%	5%	3%	6%	83%	0%
	normal	10%	10%	16%	63%	0%	0%	10%	9%	13%	68%	0%	0%	8%	10%	13%	68%	0%	0%
	BurrXII	40%	36%	15%	8%	1%	0%	41%	31%	17%	7%	3%	0%	41%	35%	15%	5%	3%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	25%	9%	43%	9%	13%	0%	29%	10%	46%	7%	7%	0%	30%	10%	45%	9%	6%	0%
χ^{2}[c]	**Weibull**	23%	24%	28%	12%	13%	0%	21%	26%	25%	16%	12%	0%	18%	40%	18%	12%	12%	0%
	gamma	16%	10%	3%	11%	60%	0%	12%	11%	8%	13%	56%	0%	9%	11%	8%	12%	60%	0%
	normal	14%	16%	16%	50%	4%	0%	12%	17%	13%	45%	13%	0%	10%	13%	16%	50%	12%	0%
	BurrXII	17%	41%	22%	17%	4%	0%	22%	34%	19%	17%	8%	0%	36%	26%	20%	11%	7%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	30%	9%	31%	10%	19%	0%	32%	13%	35%	9%	11%	0%	27%	10%	38%	15%	10%	0%
Weighted[d]	**Weibull**	20%	38%	21%	10%	11%	0%	18%	37%	25%	12%	9%	0%	16%	42%	24%	10%	7%	0%
	gamma	5%	6%	3%	17%	69%	0%	3%	5%	3%	15%	74%	0%	2%	3%	5%	11%	79%	0%
	normal	12%	9%	23%	53%	2%	0%	11%	8%	18%	55%	7%	0%	9%	9%	18%	60%	4%	0%
	BurrXII	48%	31%	12%	8%	2%	0%	44%	33%	14%	8%	1%	0%	49%	30%	12%	7%	3%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	15%	16%	41%	12%	16%	0%	24%	16%	40%	11%	8%	0%	24%	16%	42%	12%	7%	0%

* Highlighted values represent the distributional form that fits the highest percentage of farms by ranking and goodness-of-fit test.

[a] Anderson-Darling Test - (Stephens, 1974)

[b] Kolmogorov-Smirnov Test - (Chakravart et al., 1967)

[c] χ^2 Test - (Snedecor and Cochran, 1989)

[d] Weighted Test = (.334*A-D + .333*K-S +.333*χ^2)

Table 13: Goodness-of-Fit Results Grouped by Acreage: Corn Farms

		Bins (in acres)																	
		$80.6 \leq acreage < 228.8$						$228.8 \leq acreage < 364.7$						$364.7 \leq acreage \leq 2,650.4$					
		Goodness-of-Fit Rankings*																	
Test	**Distribution**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
A-D[a]	**Weibull**	10%	42%	27%	12%	9%	0%	13%	49%	26%	6%	6%	0%	13%	47%	27%	5%	8%	0%
	gamma	2%	5%	3%	17%	74%	0%	1%	3%	1%	16%	79%	0%	1%	4%	2%	14%	79%	0%
	normal	10%	8%	19%	62%	0%	0%	7%	5%	16%	72%	0%	0%	11%	4%	15%	70%	0%	0%
	BurrXII	43%	36%	15%	5%	1%	0%	42%	36%	18%	3%	1%	0%	42%	37%	15%	5%	0%	0%
	invGauss	0%	0%	0%	0%	10%	90%	0%	0%	0%	0%	10%	90%	0%	0%	0%	0%	8%	92%
	beta	34%	9%	36%	5%	6%	10%	37%	6%	39%	4%	4%	10%	33%	8%	41%	6%	5%	8%
K-S[b]	**Weibull**	16%	36%	28%	10%	10%	0%	18%	41%	28%	7%	6%	0%	18%	41%	29%	5%	8%	0%
	gamma	6%	7%	2%	9%	77%	0%	4%	5%	2%	9%	80%	0%	4%	5%	1%	10%	79%	0%
	normal	11%	10%	15%	64%	0%	0%	9%	7%	15%	70%	0%	0%	10%	8%	14%	69%	0%	0%
	BurrXII	32%	39%	18%	8%	2%	0%	32%	42%	18%	6%	1%	0%	36%	41%	16%	6%	1%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	36%	8%	36%	9%	11%	0%	37%	5%	38%	8%	12%	0%	33%	5%	40%	10%	12%	0%
χ^{2}[c]	**Weibull**	19%	26%	29%	14%	12%	0%	17%	29%	30%	13%	10%	0%	17%	32%	28%	11%	11%	0%
	gamma	11%	11%	9%	14%	55%	0%	12%	10%	7%	11%	58%	0%	12%	8%	6%	12%	62%	0%
	normal	12%	18%	16%	43%	10%	0%	15%	14%	11%	50%	9%	0%	9%	17%	15%	50%	9%	0%
	BurrXII	19%	37%	21%	17%	7%	0%	20%	38%	21%	15%	6%	0%	22%	35%	21%	17%	4%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	39%	8%	25%	12%	16%	0%	36%	8%	30%	10%	16%	0%	39%	7%	29%	10%	14%	0%
Weighted[d]	**Weibull**	18%	34%	28%	13%	7%	0%	16%	44%	25%	10%	6%	0%	18%	40%	28%	7%	7%	0%
	gamma	3%	6%	4%	14%	73%	0%	3%	3%	3%	16%	74%	0%	2%	3%	3%	16%	76%	0%
	normal	11%	8%	23%	52%	6%	0%	8%	7%	22%	57%	6%	0%	9%	7%	19%	60%	5%	0%
	BurrXII	40%	36%	14%	8%	2%	0%	43%	31%	18%	6%	1%	0%	46%	33%	13%	8%	0%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	28%	15%	32%	13%	12%	0%	29%	15%	32%	11%	13%	0%	25%	17%	36%	10%	12%	0%

* Highlighted values represent the distributional form that fits the highest percentage of farms by ranking and goodness-of-fit test.

[a] Anderson-Darling Test - (Stephens, 1974)

[b] Kolmogorov-Smirnov Test - (Chakravart et al., 1967)

[c] χ^2 Test - (Snedecor and Cochran, 1989)

[d] Weighted Test = (.334*A-D + .333*K-S +.333*χ^2)

Table 14: Goodness-of-Fit Results Grouped by Acreage: Soybean Farms

		Bins (in acres)																	
		$80.1 \leq acreage < 207.5$						$207.5 \leq acreage < 328.0$						$328.0 \leq acreage \leq 1,360.5$					
		Goodness-of-Fit Rankings*																	
Test	**Distribution**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
A-D[a]	**Weibull**	13%	45%	20%	12%	10%	0%	15%	47%	22%	6%	10%	0%	12%	47%	20%	8%	13%	0%
	gamma	2%	5%	1%	15%	76%	0%	1%	3%	2%	16%	78%	0%	2%	3%	3%	18%	74%	0%
	normal	10%	9%	18%	62%	0%	0%	9%	7%	15%	69%	0%	0%	9%	8%	18%	64%	0%	0%
	BurrXII	48%	31%	16%	4%	0%	0%	50%	33%	13%	4%	0%	0%	54%	30%	12%	4%	0%	0%
	invGauss	0%	0%	0%	0%	9%	91%	0%	0%	0%	0%	7%	93%	0%	0%	0%	0%	9%	91%
	beta	26%	10%	45%	6%	4%	9%	25%	9%	49%	5%	4%	7%	23%	11%	47%	6%	3%	9%
K-S[b]	**Weibull**	18%	36%	26%	10%	11%	0%	17%	43%	19%	9%	12%	0%	16%	41%	22%	10%	11%	0%
	gamma	6%	7%	3%	8%	77%	0%	4%	8%	2%	9%	76%	0%	4%	6%	3%	8%	79%	0%
	normal	9%	11%	15%	65%	0%	0%	11%	9%	13%	67%	0%	0%	9%	9%	15%	68%	0%	0%
	BurrXII	38%	39%	14%	8%	2%	0%	42%	30%	19%	6%	3%	0%	43%	34%	15%	6%	2%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	29%	8%	44%	10%	10%	0%	26%	10%	47%	8%	9%	0%	28%	11%	45%	8%	7%	0%
χ^2[c]	**Weibull**	20%	28%	26%	14%	13%	0%	22%	31%	23%	14%	10%	0%	19%	30%	23%	12%	15%	0%
	gamma	12%	10%	9%	13%	56%	0%	12%	12%	4%	11%	61%	0%	12%	11%	6%	12%	58%	0%
	normal	11%	18%	16%	45%	10%	0%	12%	13%	13%	52%	10%	0%	14%	15%	14%	48%	9%	0%
	BurrXII	23%	35%	19%	16%	7%	0%	25%	33%	22%	14%	7%	0%	25%	33%	21%	16%	5%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	34%	9%	30%	13%	14%	0%	29%	11%	37%	9%	13%	0%	29%	11%	35%	12%	12%	0%
Weighted[d]	**Weibull**	19%	34%	26%	12%	9%	0%	19%	41%	21%	9%	9%	0%	17%	39%	22%	12%	10%	0%
	gamma	4%	6%	3%	14%	73%	0%	3%	4%	5%	14%	74%	0%	3%	4%	3%	15%	75%	0%
	normal	11%	10%	19%	55%	4%	0%	11%	7%	19%	58%	4%	0%	10%	10%	21%	55%	5%	0%
	BurrXII	43%	35%	11%	8%	2%	0%	47%	29%	14%	7%	2%	0%	50%	31%	11%	7%	1%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	22%	14%	41%	12%	11%	0%	19%	19%	40%	11%	11%	0%	20%	16%	42%	12%	9%	0%

* Highlighted values represent the distributional form that fits the highest percentage of farms by ranking and goodness-of-fit test.

[a] Anderson-Darling Test - (Stephens, 1974)

[b] Kolmogorov-Smirnov Test - (Chakravart et al., 1967)

[c] χ^2 Test - (Snedecor and Cochran, 1989)

[d] Weighted Test = (.334*A-D + .333*K-S +.333*χ^2)

Table 15: Goodness-of-Fit Results Grouped by Expected Yield: Corn Farms

		Bins (in bushels/acre)																	
		$93.8 \le \mu < 164.6$						$164.6 \le \mu < 179.0$						$178.9 \le \mu \le 206.2$					
		Goodness-of-Fit Rankings*																	
Test	**Distribution**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
A-D[a]	**Weibull**	12%	40%	27%	11%	9%	0%	11%	49%	27%	7%	7%	0%	12%	47%	26%	6%	9%	0%
	gamma	3%	4%	3%	17%	72%	0%	1%	4%	2%	13%	80%	0%	1%	4%	2%	16%	77%	0%
	normal	14%	9%	18%	59%	0%	0%	9%	5%	15%	71%	0%	0%	9%	6%	16%	69%	0%	0%
	BurrXII	37%	38%	18%	6%	1%	0%	42%	36%	18%	4%	0%	0%	44%	35%	16%	5%	0%	0%
	invGauss	0%	0%	0%	0%	12%	88%	0%	0%	0%	0%	8%	92%	0%	0%	0%	0%	8%	92%
	beta	34%	8%	33%	8%	6%	12%	37%	7%	38%	6%	5%	8%	34%	8%	41%	4%	5%	8%
K-S[b]	**Weibull**	18%	34%	28%	10%	10%	0%	17%	40%	28%	8%	7%	0%	17%	40%	27%	6%	11%	0%
	gamma	7%	7%	2%	10%	73%	0%	4%	5%	2%	8%	81%	0%	4%	6%	2%	10%	77%	0%
	normal	14%	9%	15%	62%	0%	0%	10%	9%	14%	67%	0%	0%	10%	8%	14%	68%	0%	0%
	BurrXII	28%	42%	19%	10%	2%	0%	32%	40%	19%	7%	1%	0%	36%	39%	17%	7%	2%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	34%	8%	36%	7%	15%	0%	37%	6%	36%	9%	11%	0%	34%	7%	40%	9%	11%	0%
χ^2[c]	**Weibull**	17%	25%	30%	15%	12%	0%	18%	30%	29%	13%	10%	0%	17%	31%	27%	13%	13%	0%
	gamma	14%	12%	7%	14%	53%	0%	11%	11%	7%	11%	61%	0%	12%	11%	8%	14%	55%	0%
	normal	15%	16%	15%	43%	11%	0%	12%	15%	11%	52%	9%	0%	13%	17%	15%	45%	10%	0%
	BurrXII	17%	36%	21%	18%	7%	0%	21%	37%	22%	15%	5%	0%	22%	34%	20%	17%	7%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	36%	10%	26%	10%	18%	0%	38%	7%	31%	9%	15%	0%	36%	7%	30%	12%	15%	0%
Weighted[d]	**Weibull**	19%	33%	28%	11%	10%	0%	17%	41%	28%	9%	5%	0%	17%	41%	24%	10%	8%	0%
	gamma	5%	6%	5%	17%	67%	0%	2%	4%	3%	14%	77%	0%	3%	4%	3%	15%	74%	0%
	normal	15%	9%	20%	50%	5%	0%	9%	6%	19%	61%	5%	0%	9%	8%	23%	54%	6%	0%
	BurrXII	33%	38%	17%	11%	2%	0%	44%	33%	15%	7%	1%	0%	45%	30%	16%	8%	1%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	28%	15%	31%	11%	16%	0%	28%	16%	35%	10%	12%	0%	27%	16%	34%	12%	11%	0%

* Highlighted values represent the distributional form that fits the highest percentage of farms by ranking and goodness-of-fit test.

[a] Anderson-Darling Test - (Stephens, 1974)

[b] Kolmogorov-Smirnov Test - (Chakravart et al., 1967)

[c] χ^2 Test - (Snedecor and Cochran, 1989)

[d] Weighted Test = (.334*A-D + .333*K-S +.333*χ^2)

Table 16: Goodness-of-Fit Results Grouped by Expected Yield: Soybean Farms

		Bins (in bushels/acre)																	
		$31.7 \leq \mu < 49.8$						$49.8 \leq \mu < 53.9$						$53.9 \leq \mu \leq 65.8$					
		Goodness-of-Fit Rankings*																	
Test	**Distribution**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
A-D[a]	**Weibull**	13%	42%	23%	11%	12%	0%	14%	50%	20%	9%	8%	0%	12%	48%	18%	8%	14%	0%
	gamma	4%	4%	2%	18%	72%	0%	1%	4%	1%	14%	81%	0%	1%	5%	2%	17%	75%	0%
	normal	12%	12%	19%	57%	0%	0%	7%	6%	16%	71%	0%	0%	10%	7%	16%	67%	0%	0%
	BurrXII	43%	34%	17%	6%	0%	0%	51%	33%	12%	3%	0%	0%	57%	26%	13%	4%	0%	0%
	invGauss	0%	0%	0%	0%	10%	90%	0%	0%	0%	0%	8%	92%	0%	0%	0%	0%	8%	92%
	beta	29%	8%	39%	9%	6%	10%	27%	7%	50%	4%	3%	8%	19%	14%	51%	5%	3%	8%
K-S[b]	**Weibull**	18%	36%	23%	11%	12%	0%	16%	41%	25%	10%	8%	0%	17%	42%	19%	8%	14%	0%
	gamma	7%	7%	3%	9%	74%	0%	4%	6%	2%	5%	83%	0%	4%	8%	3%	11%	74%	0%
	normal	10%	12%	16%	61%	0%	0%	8%	7%	13%	72%	0%	0%	10%	9%	14%	67%	0%	0%
	BurrXII	35%	34%	20%	9%	3%	0%	40%	37%	15%	6%	2%	0%	47%	31%	13%	6%	2%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	30%	10%	38%	10%	12%	0%	33%	9%	45%	8%	6%	0%	22%	10%	51%	8%	9%	0%
χ^2[c]	**Weibull**	23%	26%	23%	15%	13%	0%	19%	33%	26%	12%	10%	0%	20%	30%	22%	13%	14%	0%
	gamma	16%	10%	7%	10%	56%	0%	9%	10%	6%	11%	64%	0%	13%	12%	6%	14%	55%	0%
	normal	13%	18%	15%	45%	8%	0%	10%	12%	14%	54%	10%	0%	14%	15%	16%	45%	10%	0%
	BurrXII	21%	34%	22%	17%	7%	0%	25%	36%	20%	13%	6%	0%	27%	32%	18%	16%	7%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	28%	12%	32%	12%	16%	0%	37%	9%	34%	10%	9%	0%	26%	11%	38%	11%	14%	0%
Weighted[d]	**Weibull**	21%	32%	26%	12%	10%	0%	17%	43%	23%	10%	7%	0%	17%	41%	20%	11%	11%	0%
	gamma	4%	6%	5%	17%	68%	0%	2%	4%	2%	11%	80%	0%	3%	5%	4%	15%	72%	0%
	normal	12%	13%	21%	49%	4%	0%	8%	6%	18%	64%	4%	0%	12%	8%	20%	54%	5%	0%
	BurrXII	40%	33%	16%	9%	2%	0%	49%	32%	11%	6%	2%	0%	52%	29%	11%	7%	2%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	23%	17%	32%	13%	15%	0%	24%	15%	45%	9%	7%	0%	16%	17%	45%	12%	9%	0%

* Highlighted values represent the distributional form that fits the highest percentage of farms by ranking and goodness-of-fit test.

[a] Anderson-Darling Test - (Stephens, 1974)

[b] Kolmogorov-Smirnov Test - (Chakravart et al., 1967)

[c] χ^2 Test - (Snedecor and Cochran, 1989)

[d] Weighted Test = (.334*A-D + .333*K-S +.333*χ^2)

Table 17: Goodness-of-Fit Results Grouped by Standard Deviation: Corn Farms

		Bins (in bushels/acre)																	
		$13.3 \leq \sigma < 23.9$						$23.9 \leq \sigma < 27.8$						$27.8 \leq \sigma \leq 43.5$					
		Goodness-of-Fit Rankings*																	
Test	**Distribution**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
A-D[a]	**Weibull**	11%	42%	24%	10%	14%	0%	11%	48%	25%	8%	8%	0%	14%	46%	31%	6%	4%	0%
	gamma	3%	5%	2%	14%	76%	0%	1%	4%	2%	16%	76%	0%	1%	2%	2%	17%	77%	0%
	normal	11%	9%	13%	67%	0%	0%	12%	7%	15%	66%	0%	0%	9%	4%	22%	65%	0%	0%
	BurrXII	43%	32%	19%	6%	0%	0%	42%	35%	18%	4%	1%	0%	38%	42%	16%	4%	0%	0%
	invGauss	0%	0%	0%	0%	4%	96%	0%	0%	0%	0%	9%	91%	0%	0%	0%	0%	14%	86%
	beta	32%	12%	43%	4%	5%	4%	33%	6%	40%	6%	6%	9%	39%	5%	29%	8%	5%	14%
K-S[b]	**Weibull**	18%	35%	25%	9%	13%	0%	16%	38%	27%	9%	10%	0%	17%	41%	31%	6%	4%	0%
	gamma	7%	7%	3%	10%	73%	0%	6%	7%	2%	9%	76%	0%	3%	4%	1%	10%	82%	0%
	normal	11%	10%	14%	64%	0%	0%	12%	9%	14%	65%	0%	0%	9%	7%	15%	69%	0%	0%
	BurrXII	32%	38%	18%	9%	2%	0%	34%	38%	17%	10%	2%	0%	30%	45%	20%	5%	1%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	31%	10%	40%	8%	12%	0%	33%	7%	40%	7%	12%	0%	41%	3%	33%	10%	13%	0%
χ^2[c]	**Weibull**	21%	28%	23%	12%	15%	0%	16%	29%	29%	14%	12%	0%	14%	29%	34%	14%	8%	0%
	gamma	13%	12%	7%	13%	55%	0%	13%	12%	8%	13%	54%	0%	11%	10%	7%	11%	60%	0%
	normal	13%	17%	14%	47%	8%	0%	16%	15%	14%	44%	11%	0%	11%	16%	14%	49%	10%	0%
	BurrXII	21%	34%	19%	18%	8%	0%	22%	36%	21%	16%	5%	0%	17%	38%	23%	16%	5%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	31%	9%	37%	9%	14%	0%	33%	7%	29%	12%	18%	0%	46%	7%	22%	9%	16%	0%
Weighted[d]	**Weibull**	18%	36%	22%	13%	11%	0%	16%	39%	27%	10%	8%	0%	18%	39%	31%	8%	3%	0%
	gamma	4%	5%	4%	13%	73%	0%	3%	5%	5%	18%	69%	0%	2%	3%	3%	16%	76%	0%
	normal	11%	11%	20%	54%	3%	0%	13%	8%	20%	52%	7%	0%	9%	5%	21%	58%	5%	0%
	BurrXII	40%	30%	19%	11%	1%	0%	41%	35%	13%	9%	2%	0%	41%	36%	16%	6%	1%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	26%	18%	35%	10%	11%	0%	27%	12%	36%	11%	14%	0%	29%	16%	29%	11%	14%	0%

* Highlighted values represent the distributional form that fits the highest percentage of farms by ranking and goodness-of-fit test.

[a] Anderson-Darling Test - (Stephens, 1974)

[b] Kolmogorov-Smirnov Test - (Chakravart et al., 1967)

[c] χ^2 Test - (Snedecor and Cochran, 1989)

[d] Weighted Test = (.334*A-D + .333*K-S +.333*χ^2)

Table 18: Goodness-of-Fit Results Grouped by Standard Deviation: Soybean Farms

		Bins (in bushels/acre)																	
		$3.3 \leq \sigma < 6.4$						$6.4 \leq \sigma < 7.7$						$7.7 \leq \sigma \leq 12.8$					
		Goodness-of-Fit Rankings[*]																	
Test	**Distribution**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
A-D[a]	**Weibull**	12%	43%	18%	10%	18%	0%	14%	50%	18%	10%	8%	0%	13%	47%	25%	8%	7%	0%
	gamma	2%	7%	3%	15%	73%	0%	2%	3%	1%	15%	78%	0%	2%	2%	1%	18%	77%	0%
	normal	13%	7%	16%	65%	0%	0%	9%	9%	15%	67%	0%	0%	7%	9%	21%	62%	0%	0%
	BurrXII	54%	26%	13%	6%	0%	0%	51%	30%	16%	4%	1%	0%	47%	37%	13%	3%	0%	0%
	invGauss	0%	0%	0%	0%	4%	96%	0%	0%	0%	0%	10%	90%	0%	0%	0%	0%	11%	89%
	beta	19%	18%	50%	4%	4%	4%	25%	9%	50%	4%	3%	10%	31%	4%	40%	10%	4%	11%
K-S[b]	**Weibull**	16%	37%	18%	10%	19%	0%	17%	42%	22%	11%	7%	0%	17%	40%	26%	8%	8%	0%
	gamma	6%	10%	4%	10%	70%	0%	3%	6%	2%	9%	80%	0%	6%	5%	2%	6%	82%	0%
	normal	13%	10%	14%	63%	0%	0%	9%	9%	14%	68%	0%	0%	7%	10%	15%	69%	0%	0%
	BurrXII	43%	28%	16%	9%	3%	0%	42%	34%	15%	6%	3%	0%	37%	39%	17%	6%	1%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	22%	14%	48%	8%	8%	0%	28%	9%	46%	6%	10%	0%	34%	6%	41%	11%	9%	0%
χ^2[c]	**Weibull**	24%	26%	21%	11%	18%	0%	19%	32%	23%	16%	11%	0%	21%	30%	27%	14%	9%	0%
	gamma	13%	13%	7%	11%	56%	0%	12%	10%	8%	12%	58%	0%	12%	9%	5%	12%	62%	0%
	normal	13%	14%	15%	49%	9%	0%	13%	16%	14%	49%	9%	0%	11%	16%	16%	47%	10%	0%
	BurrXII	23%	33%	20%	18%	6%	0%	28%	31%	20%	14%	6%	0%	22%	37%	21%	13%	7%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	26%	13%	37%	11%	12%	0%	29%	11%	36%	9%	16%	0%	35%	8%	31%	14%	12%	0%
Weighted[d]	**Weibull**	18%	35%	20%	13%	14%	0%	17%	42%	22%	10%	8%	0%	19%	39%	28%	8%	7%	0%
	gamma	4%	6%	5%	14%	71%	0%	3%	3%	3%	16%	74%	0%	3%	5%	3%	14%	76%	0%
	normal	14%	10%	19%	52%	5%	0%	10%	8%	21%	56%	4%	0%	8%	9%	19%	59%	4%	0%
	BurrXII	46%	30%	12%	9%	2%	0%	48%	30%	13%	7%	2%	0%	46%	34%	13%	6%	1%	0%
	invGauss	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%
	beta	17%	19%	44%	11%	9%	0%	21%	16%	41%	11%	11%	0%	24%	14%	37%	13%	12%	0%

[*] Highlighted values represent the distributional form that fits the highest percentage of farms by ranking and goodness-of-fit test.

[a] Anderson-Darling Test - (Stephens, 1974)

[b] Kolmogorov-Smirnov Test - (Chakravart et al., 1967)

[c] χ^2 Test - (Snedecor and Cochran, 1989)

[d] Weighted Test = (.334*A-D + .333*K-S +.333*χ^2)

Table 19: Insurance Rate Statistics: Illinois by Districts for FBFM Corn Farms at Three Coverage Levels.

		Illinois Districts														
		NW			**NE**			**West**			**Central**			**East**		
		Coverage Levels[*+]														
Statistics	**Distribution**	**85%**	**75%**	**65%**	**85%**	**75%**	**65%**	**85%**	**75%**	**65%**	**85%**	**75%**	**65%**	**85%**	**75%**	**65%**
Average[a]	**Empirical**[b]	2.44	1.01	0.38	2.66	1.05	0.36	3.39	1.50	0.51	2.84	1.13	0.37	3.84	1.73	0.68
	Weibull	1.97	0.63	0.18	2.30	0.78	0.24	2.98	1.09	0.35	2.30	0.74	0.21	2.75	0.99	0.31
	normal	1.84	0.48	0.11	2.12	0.62	0.16	2.80	0.86	0.23	2.13	0.55	0.12	2.91	0.94	0.26
	Burr XII	1.86	0.58	0.16	2.17	0.73	0.22	2.82	1.00	0.31	2.17	0.68	0.19	2.70	0.96	0.30
	beta	2.60	0.99	0.36	2.83	1.14	0.44	3.52	1.42	0.51	2.93	1.12	0.40	3.87	1.72	0.71
	Mix2Norm	2.11	0.73	0.22	2.56	0.95	0.32	3.20	1.25	0.39	2.67	0.96	0.29	3.71	1.58	0.57
	kernel	2.82	1.18	0.46	3.10	1.27	0.47	3.94	1.74	0.67	3.24	1.32	0.47	4.28	2.00	0.83
Bias (%)[c]	**Weibull**	-19.2%	-37.8%	-53.5%	-13.4%	-25.6%	-33.6%	-11.9%	-27.1%	-31.1%	-19.2%	-34.3%	-43.8%	-28.4%	-43.0%	-54.5%
	normal	-24.8%	-52.8%	-70.8%	-20.4%	-41.4%	-54.4%	-17.3%	-42.2%	-55.1%	-25.1%	-51.0%	-67.0%	-24.4%	-45.9%	-61.7%
	Burr XII	-23.9%	-43.0%	-58.5%	-18.3%	-31.1%	-39.4%	-16.7%	-32.9%	-37.9%	-23.6%	-39.6%	-49.7%	-29.7%	-44.5%	-55.9%
	beta	**6.4%**	-2.2%	-7.4%	**6.4%**	**7.9%**	**21.5%**	**3.9%**	-5.2%	**1.5%**	**3.2%**	-1.0%	**6.7%**	**0.8%**	-0.7%	**3.8%**
	Mix2Norm	-13.8%	-28.2%	-42.9%	-3.6%	-9.8%	-11.8%	-5.3%	-16.6%	-23.6%	-6.2%	-15.0%	-20.7%	-3.6%	-8.5%	-15.4%
	kernel	**15.3%**	**16.3%**	**20.2%**	**16.5%**	**20.3%**	**31.7%**	**16.4%**	**16.7%**	**32.4%**	**13.8%**	**16.9%**	**25.8%**	**11.5%**	**15.6%**	**22.4%**
RMSE (%)[d]	**Weibull**	41.6%	81.3%	137.5%	37.0%	73.0%	126.6%	32.1%	56.6%	92.4%	38.0%	71.2%	124.7%	39.6%	67.9%	105.5%
	normal	32.0%	77.2%	135.2%	27.9%	67.3%	121.0%	23.9%	55.6%	93.4%	30.7%	69.6%	124.2%	28.8%	60.4%	100.2%
	Burr XII	40.1%	80.5%	136.9%	35.2%	72.1%	126.2%	30.2%	56.4%	92.1%	36.6%	70.0%	123.3%	39.4%	67.7%	105.5%
	beta	29.7%	62.6%	118.7%	27.0%	57.1%	122.4%	19.1%	31.9%	63.1%	25.2%	54.8%	111.1%	20.3%	41.6%	77.0%
	Mix2Norm	36.3%	64.5%	110.0%	13.7%	27.4%	49.4%	18.3%	35.2%	65.3%	21.2%	38.7%	72.8%	14.7%	25.6%	46.9%
	kernel	18.4%	24.1%	34.8%	18.9%	27.2%	50.2%	18.1%	20.5%	46.5%	16.2%	22.1%	37.6%	13.7%	20.0%	31.3%

Continued on next page

Table 19 – continued from previous page

		Illinois Districts														
		WSW			**ESE**			**SW**			**SE**			**Total**		
		Coverage Levels[*+]														
Statistics	**Distribution**	**85%**	**75%**	**65%**	**85%**	**75%**	**65%**	**85%**	**75%**	**65%**	**85%**	**75%**	**65%**	**85%**	**75%**	**65%**
Average[a]	**Empirical[b]**	2.34	0.81	0.21	3.33	1.27	0.42	3.70	1.56	0.56	3.14	1.23	0.42	3.00	1.23	0.43
	Weibull	2.13	0.69	0.20	3.21	1.29	0.46	3.57	1.55	0.60	3.02	1.22	0.44	2.53	0.90	0.29
	normal	1.85	0.47	0.10	2.97	1.05	0.33	3.36	1.33	0.47	2.82	1.02	0.32	2.39	0.73	0.20
	Burr XII	2.02	0.64	0.18	3.06	1.21	0.42	3.39	1.45	0.55	2.88	1.14	0.40	2.41	0.84	0.26
	beta	2.67	1.00	0.35	3.77	1.70	0.71	3.88	1.78	0.75	3.37	1.48	0.60	3.18	1.31	0.50
	Mix2Norm	2.13	0.67	0.16	3.26	1.19	0.38	3.62	1.47	0.51	3.07	1.12	0.35	2.83	1.06	0.34
	kernel	2.77	1.01	0.31	3.98	1.68	0.63	4.33	1.99	0.81	3.75	1.61	0.62	3.46	1.48	0.56
Bias (%)[c]	**Weibull**	-9.1%	-15.2%	-7.7%	-3.5%	**1.2%**	**10.2%**	-3.5%	-0.2%	**7.0%**	-3.6%	-0.7%	**3.8%**	-15.7%	-26.5%	-33.3%
	normal	-20.7%	-42.1%	-51.0%	-10.6%	-17.3%	-20.0%	-9.2%	-14.6%	-16.8%	-9.9%	-17.6%	-23.5%	-20.3%	-40.6%	-53.5%
	Burr XII	-13.5%	-21.6%	-16.8%	-7.9%	-5.3%	**1.2%**	-8.2%	-7.1%	-2.5%	-8.2%	-7.5%	-5.1%	-19.7%	-31.4%	-39.0%
	beta	**14.3%**	**22.8%**	**63.1%**	**13.5%**	**33.3%**	**70.5%**	**4.9%**	**14.1%**	**32.8%**	**7.5%**	**20.0%**	**40.6%**	**5.8%**	**6.6%**	**17.0%**
	Mix2Norm	-8.8%	-17.0%	-26.0%	-2.1%	-6.6%	-9.6%	-2.0%	-5.5%	-9.3%	-2.1%	-9.2%	-17.1%	-5.8%	-13.3%	-20.9%
	kernel	**18.4%**	**24.8%**	**47.3%**	**19.7%**	**31.9%**	**50.4%**	**17.3%**	**28.2%**	**44.5%**	**19.7%**	**30.9%**	**46.5%**	**15.4%**	**20.4%**	**30.6%**
RMSE (%)[d]	**Weibull**	38.2%	79.9%	179.8%	26.6%	52.7%	101.4%	21.0%	42.8%	86.7%	26.3%	51.4%	104.2%	36.0%	68.0%	118.7%
	normal	30.2%	74.1%	169.3%	18.4%	42.2%	89.0%	15.2%	34.6%	74.4%	18.2%	42.6%	94.2%	27.3%	62.6%	113.7%
	Burr XII	36.2%	77.8%	175.6%	24.1%	49.3%	96.6%	19.3%	40.5%	83.4%	23.1%	49.6%	103.2%	34.6%	66.9%	117.5%
	beta	35.1%	73.5%	189.2%	28.0%	66.8%	148.1%	18.2%	42.0%	90.8%	22.3%	51.6%	110.3%	25.3%	54.1%	109.8%
	Mix2Norm	27.6%	41.6%	154.6%	11.4%	22.7%	50.5%	8.7%	19.4%	42.8%	11.6%	26.1%	58.3%	20.5%	36.7%	71.3%
	kernel	21.2%	31.5%	71.5%	21.4%	37.9%	72.3%	19.0%	32.3%	55.7%	21.7%	37.2%	64.1%	18.1%	27.1%	47.1%

[*] The highlighted values represent the distributional form with the lowest absolute bias or efficiency value by district and coverage level.

[+] The bolded values represent bias values for which the fitted rate is greater than the empirical/burn rate.

[a] Expected value of fitted distributional rates over all farms by district and coverage level.

[b] Empirical/burn rate [max(0,Yield Guarantee-Estimated Rate)] for all farms by district and coverage level.

[c] Bias (%) is expected value of [empirical/burn rate - fitted distributional rate] for all farms divided by empirical/burn rate by district and coverage level.

[d] RMSE (%) is the expected value of the RMSE for all farms divided by empirical/burn rate by district and coverage level.

Table 20: Insurance Rate Statistics: Illinois by Districts for FBFM Soybean Farms at Three Coverage Levels.

		Illinois Districts														
		NW			**NE**			**West**			**Central**			**East**		
		Coverage Levels[*+]														
Statistics	**Distribution**	**85%**	**75%**	**65%**	**85%**	**75%**	**65%**	**85%**	**75%**	**65%**	**85%**	**75%**	**65%**	**85%**	**75%**	**65%**
Average[a]	**Empirical**[b]	0.83	0.36	0.13	0.66	0.26	0.09	0.54	0.16	0.04	0.66	0.25	0.07	0.90	0.40	0.15
	Weibull	0.71	0.24	0.08	0.57	0.18	0.05	0.56	0.17	0.05	0.51	0.15	0.04	0.70	0.24	0.07
	normal	0.64	0.18	0.05	0.50	0.13	0.03	0.43	0.10	0.02	0.44	0.10	0.02	0.67	0.19	0.05
	Burr XII	0.63	0.20	0.06	0.52	0.16	0.05	0.49	0.14	0.04	0.45	0.13	0.03	0.64	0.21	0.06
	beta	0.83	0.31	0.11	0.66	0.23	0.07	0.60	0.21	0.07	0.59	0.19	0.06	0.89	0.35	0.13
	Mix2Norm	0.75	0.30	0.11	0.61	0.23	0.08	0.51	0.15	0.04	0.62	0.22	0.06	0.86	0.36	0.14
	kernel	0.94	0.41	0.16	0.76	0.31	0.11	0.64	0.20	0.05	0.73	0.28	0.09	1.01	0.45	0.18
Bias (%)[c]	**Weibull**	-14.3%	-32.7%	-40.2%	-13.6%	-28.3%	-39.7%	**3.8%**	**11.1%**	**35.9%**	-22.1%	-38.4%	-41.3%	-22.3%	-40.6%	-54.7%
	normal	-22.9%	-49.8%	-62.3%	-25.3%	-50.3%	-64.2%	-20.6%	-36.4%	-39.1%	-33.3%	-60.3%	-70.7%	-25.2%	-51.7%	-69.7%
	Burr XII	-23.9%	-43.8%	-53.3%	-21.3%	-37.0%	-48.0%	-9.6%	-9.6%	**3.4%**	-31.4%	-48.8%	-53.5%	-28.6%	-47.7%	-61.8%
	beta	**0.3%**	-13.7%	-12.7%	-0.2%	-10.4%	-15.8%	**12.6%**	**32.2%**	**92.1%**	-10.4%	-24.3%	-19.5%	-0.8%	-11.5%	-16.6%
	Mix2Norm	-9.0%	-17.3%	-13.0%	-8.7%	-11.4%	-10.6%	-5.4%	-4.7%	**13.9%**	-6.4%	-11.1%	-9.9%	-4.1%	-9.2%	-11.2%
	kernel	**13.5%**	**13.4%**	**23.8%**	**15.0%**	**19.8%**	**24.3%**	**18.9%**	**29.1%**	**50.5%**	**11.2%**	**14.1%**	**26.6%**	**12.1%**	**13.5%**	**17.8%**
RMSE (%)[d]	**Weibull**	40.9%	72.3%	130.0%	42.2%	85.7%	137.7%	42.2%	103.2%	227.9%	47.6%	89.7%	161.2%	41.3%	75.3%	117.9%
	normal	30.7%	68.6%	125.2%	33.9%	83.0%	138.2%	30.5%	82.9%	184.6%	41.6%	90.1%	161.1%	31.1%	71.0%	116.7%
	Burr XII	37.6%	72.9%	130.7%	40.9%	85.4%	137.1%	35.4%	86.6%	198.6%	47.9%	90.8%	160.9%	40.7%	75.6%	118.3%
	beta	29.9%	55.7%	104.9%	31.9%	70.0%	117.5%	42.1%	101.8%	281.7%	30.1%	69.9%	140.1%	25.9%	51.8%	87.8%
	Mix2Norm	28.8%	44.7%	76.5%	22.8%	35.6%	54.5%	16.3%	29.0%	64.4%	21.0%	38.6%	83.4%	16.9%	29.0%	45.6%
	kernel	17.1%	20.4%	38.5%	19.1%	32.3%	50.9%	22.4%	40.9%	85.7%	15.9%	23.8%	51.4%	15.0%	18.2%	26.5%

Continued on next page

Table 20 – continued from previous page

		Illinois Districts														
		WSW			**ESE**			**SW**			**SE**			**Total**		
		Coverage Levels*+														
Statistics	**Distribution**	**85%**	**75%**	**65%**	**85%**	**75%**	**65%**	**85%**	**75%**	**65%**	**85%**	**75%**	**65%**	**85%**	**75%**	**65%**
Average[a]	**Empirical**[b]	0.58	0.18	0.04	0.70	0.25	0.07	1.04	0.43	0.16	0.85	0.33	0.10	0.74	0.29	0.09
	Weibull	0.54	0.17	0.05	0.70	0.25	0.08	1.07	0.46	0.18	0.87	0.35	0.13	0.65	0.22	0.07
	normal	0.44	0.11	0.02	0.60	0.18	0.05	0.97	0.37	0.13	0.75	0.26	0.09	0.57	0.16	0.04
	Burr XII	0.48	0.14	0.04	0.64	0.22	0.07	0.97	0.39	0.14	0.76	0.29	0.10	0.58	0.19	0.06
	beta	0.58	0.19	0.06	0.78	0.30	0.11	1.09	0.47	0.18	0.94	0.40	0.17	0.74	0.27	0.10
	Mix2Norm	0.55	0.16	0.04	0.67	0.23	0.07	1.02	0.40	0.14	0.84	0.31	0.10	0.70	0.26	0.08
	kernel	0.67	0.23	0.06	0.84	0.32	0.10	1.23	0.55	0.22	1.00	0.42	0.15	0.85	0.34	0.12
Bias (%)[c]	**Weibull**	-6.2%	-3.7%	**20.1%**	-0.2%	**0.8%**	**22.4%**	**3.7%**	**7.4%**	**13.6%**	**1.7%**	**6.2%**	**23.1%**	-12.8%	-24.2%	-27.3%
	normal	-24.5%	-40.1%	-40.0%	-14.4%	-28.4%	-26.1%	-6.8%	-14.5%	-19.5%	-11.5%	-19.3%	-16.6%	-23.3%	-45.3%	-55.7%
	Burr XII	-16.5%	-19.8%	-8.3%	-8.9%	-12.3%	**1.4%**	-6.7%	-8.6%	-8.8%	-10.3%	-10.9%	-0.9%	-21.8%	-35.4%	-41.1%
	beta	-0.2%	**8.2%**	**50.9%**	**10.6%**	**19.6%**	**61.4%**	**5.0%**	**7.9%**	**13.9%**	**9.7%**	**23.1%**	**61.8%**	-0.2%	-6.1%	**1.2%**
	Mix2Norm	-5.0%	-9.2%	**1.6%**	-3.9%	-9.5%	-2.8%	-1.6%	-7.0%	-12.0%	-1.7%	-4.0%	-2.3%	-5.5%	-10.8%	-9.6%
	kernel	**16.6%**	**28.4%**	**50.6%**	**19.6%**	**26.9%**	**53.7%**	**18.6%**	**28.2%**	**42.0%**	**17.2%**	**28.2%**	**51.0%**	**14.5%**	**18.6%**	**29.9%**
RMSE (%)[d]	**Weibull**	45.8%	107.2%	246.7%	36.5%	72.7%	162.1%	26.2%	49.8%	96.8%	35.2%	63.8%	146.7%	41.5%	79.6%	144.0%
	normal	36.6%	92.5%	208.3%	25.4%	60.9%	139.1%	16.0%	37.8%	82.6%	20.0%	46.1%	116.0%	32.4%	74.3%	136.9%
	Burr XII	42.1%	97.6%	206.0%	29.5%	61.9%	144.3%	18.8%	41.6%	88.7%	24.2%	52.5%	132.4%	39.0%	77.5%	140.2%
	beta	35.5%	95.9%	258.1%	32.7%	76.6%	195.4%	18.2%	40.1%	86.0%	33.1%	70.5%	174.1%	29.8%	64.3%	128.6%
	Mix2Norm	13.4%	29.9%	71.1%	15.0%	28.2%	63.0%	8.0%	18.7%	49.2%	7.6%	21.8%	50.1%	19.3%	34.2%	68.0%
	kernel	21.1%	42.7%	95.8%	23.2%	38.6%	92.1%	20.5%	33.2%	55.8%	19.3%	37.2%	80.1%	18.5%	28.1%	52.7%

* The highlighted values represent the distributional form with the lowest absolute bias or efficiency value by district and coverage level.

+The bolded values represent bias values for which the fitted rate is greater than the empirical/burn rate.

[a] Expected value of fitted distributional rates over all farms by district and coverage level.

[b] Empirical/burn rate [max(0,Yield Guarantee-Estimated Rate)] for all farms by district and coverage level.

[c] Bias (%) is expected value of [empirical/burn rate - fitted distributional rate] for all farms divided by empirical/burn rate by district and coverage level.

[d] RMSE (%) is the expected value of the RMSE for all farms divided by empirical/burn rate by district and coverage level.

Table 21: Out-of-Sample Rate Simulation Analysis: μ=160; σ=20

Data Generating Process:		***Weibull***			***beta***		
		Coverage Level^{*+}			**Coverage Level**		
Statistic	**Distribution**	85%	75%	65%	85%	75%	65%
	True Theoretical	1.59	0.43	0.10	1.42	0.30	0.04
n=10							
Average	**Empirical**	1.60	0.44	0.10	1.42	0.30	0.04
	Beta	2.80	1.39	0.70	2.81	1.39	0.70
	Weibull	1.61	0.49	0.13	1.67	0.51	0.14
	Mix2Norm	1.30	0.32	0.07	1.19	0.23	0.04
Bias (%)	**Empirical**	**0.7%**	**2.4%**	**2.4%**	-0.2%	-0.7%	**1.6%**
	Beta	**76.6%**	**222.2%**	**631.5%**	**97.3%**	**359.2%**	**1502.8%**
	Weibull	**1.2%**	**13.6%**	**38.9%**	**17.6%**	**68.6%**	**211.6%**
	Mix2Norm	-18.2%	-25.3%	-23.8%	-16.4%	-24.7%	-10.9%
RMSE (%)	**Empirical**	120.8%	223.8%	437.8%	115.3%	233.1%	592.3%
	Beta	214.6%	561.5%	1706.5%	242.0%	815.2%	3741.4%
	Weibull	90.4%	138.8%	227.8%	97.9%	197.6%	498.2%
	Mix2Norm	117.1%	184.6%	307.4%	111.2%	180.9%	382.7%
n=15							
Average	**Empirical**	1.58	0.43	0.10	1.42	0.30	0.04
	Beta	2.48	1.07	0.47	2.36	0.98	0.41
	Weibull	1.60	0.47	0.12	1.69	0.50	0.13
	Mix2Norm	1.38	0.33	0.07	1.28	0.25	0.04
Bias (%)	**Empirical**	-0.4%	**0.9%**	**5.8%**	**0.0%**	-0.8%	-4.7%
	Beta	**56.1%**	**148.7%**	**392.6%**	**65.6%**	**223.8%**	**839.5%**
	Weibull	**0.9%**	**9.3%**	**25.7%**	**18.7%**	**66.5%**	**195.4%**
	Mix2Norm	-13.2%	-23.7%	-25.3%	-10.3%	-17.8%	-4.7%
RMSE (%)	**Empirical**	96.5%	179.6%	368.0%	92.4%	184.5%	430.3%
	Beta	155.4%	378.0%	1051.5%	158.2%	487.9%	1998.2%
	Weibull	72.1%	105.3%	160.1%	80.5%	160.6%	389.7%
	Mix2Norm	96.2%	151.4%	252.2%	91.3%	151.6%	291.2%
n=20							
Average	**Empirical**	1.58	0.42	0.09	1.43	0.30	0.05
	Beta	2.12	0.80	0.30	2.05	0.74	0.27
	Weibull	1.60	0.46	0.11	1.70	0.50	0.13
	Mix2Norm	1.44	0.34	0.07	1.33	0.27	0.04
Bias (%)	**Empirical**	-0.7%	-2.1%	-5.3%	**0.5%**	**0.4%**	**7.2%**
	Beta	**33.4%**	**85.7%**	**214.3%**	**44.0%**	**146.1%**	**512.8%**
	Weibull	**0.8%**	**7.2%**	**19.8%**	**19.1%**	**65.2%**	**187.4%**
	Mix2Norm	-9.3%	-20.6%	-23.8%	-6.4%	-12.2%	**0.7%**
RMSE (%)	**Empirical**	83.3%	150.2%	293.4%	82.7%	166.7%	414.4%
	Beta	112.1%	248.6%	623.6%	118.9%	341.7%	1303.9%
	Weibull	62.0%	88.9%	131.5%	71.9%	144.3%	347.1%
	Mix2Norm	84.4%	132.2%	221.8%	82.4%	142.0%	286.0%
n=30							
Average	**Empirical**	1.59	0.44	0.10	1.42	0.30	0.04
	Beta	1.87	0.60	0.18	1.76	0.53	0.15
	Weibull	1.59	0.45	0.11	1.70	0.49	0.12
	Mix2Norm	1.50	0.37	0.08	1.36	0.27	0.04
Bias (%)	**Empirical**	**0.5%**	**1.5%**	**4.1%**	-0.1%	-0.3%	-2.4%
	Beta	**17.8%**	**40.3%**	**93.2%**	**23.3%**	**75.0%**	**238.9%**
	Weibull	-0.1%	**3.9%**	**12.1%**	**19.4%**	**63.1%**	**175.5%**
	Mix2Norm	-5.5%	-13.2%	-16.4%	-4.5%	-10.3%	**0.5%**
RMSE (%)	**Empirical**	67.4%	125.5%	251.6%	65.7%	132.7%	319.5%
	Beta	79.5%	158.6%	360.3%	77.8%	195.3%	641.1%
	Weibull	50.0%	70.6%	101.1%	58.8%	119.0%	281.5%
	Mix2Norm	69.3%	114.8%	188.5%	66.0%	116.5%	237.2%

* The highlighted values represent the distributional form with the lowest absolute bias or efficiency value by district and coverage level.

+ The bolded values represent bias values for which the fitted rate is greater than the empirical/burn rate.

Table 22: Out-of-Sample Rate Simulation Analysis: μ=160; σ=30

Data Generating Process:		***Weibull***			***beta***		
		Coverage Level^{*+}			**Coverage Level**		
Statistic	**Distribution**	85%	75%	65%	85%	75%	65%
	True Theoretical	4.10	1.72	0.62	3.99	1.54	0.48
n=10							
Average	**Empirical**	4.08	1.71	0.62	4.00	1.53	0.47
	Beta	5.55	3.23	1.84	5.63	3.27	1.86
	Weibull	4.00	1.75	0.69	4.06	1.77	0.69
	Mix2Norm	3.62	1.39	0.46	3.67	1.31	0.39
Bias (%)	**Empirical**	-0.4%	-0.6%	-0.8%	**0.2%**	-0.8%	-1.8%
	Beta	**35.5%**	**88.1%**	**196.2%**	**41.1%**	**112.5%**	**289.2%**
	Weibull	-2.3%	**1.9%**	**10.8%**	**1.8%**	**14.9%**	**44.0%**
	Mix2Norm	-11.6%	-19.0%	-25.5%	-8.1%	-14.7%	-17.9%
RMSE (%)	**Empirical**	85.0%	125.8%	196.1%	81.8%	124.6%	205.0%
	Beta	124.3%	230.7%	468.6%	131.6%	271.1%	650.4%
	Weibull	71.3%	94.0%	127.9%	69.9%	100.0%	159.9%
	Mix2Norm	87.1%	117.9%	157.5%	84.2%	117.5%	166.3%
n=15							
Average	**Empirical**	4.11	1.71	0.61	4.00	1.55	0.48
	Beta	5.27	2.81	1.46	5.25	2.79	1.43
	Weibull	4.03	1.74	0.67	4.10	1.77	0.68
	Mix2Norm	3.88	1.53	0.52	3.81	1.39	0.43
Bias (%)	**Empirical**	**0.4%**	-0.1%	-1.4%	**0.3%**	**1.0%**	**0.1%**
	Beta	**28.6%**	**63.9%**	**134.5%**	**31.7%**	**81.0%**	**198.6%**
	Weibull	-1.7%	**1.2%**	**7.3%**	**2.8%**	**15.0%**	**41.4%**
	Mix2Norm	-5.2%	-11.0%	-16.5%	-4.6%	-9.4%	-10.5%
RMSE (%)	**Empirical**	70.1%	102.8%	159.0%	67.0%	102.4%	173.9%
	Beta	98.0%	176.6%	344.7%	97.7%	194.6%	445.6%
	Weibull	58.2%	75.6%	99.8%	57.2%	81.8%	129.8%
	Mix2Norm	72.2%	100.2%	137.8%	68.9%	99.2%	149.8%
n=20							
Average	**Empirical**	4.10	1.72	0.63	4.01	1.55	0.48
	Beta	4.93	2.44	1.15	4.83	2.35	1.08
	Weibull	4.04	1.73	0.65	4.12	1.77	0.67
	Mix2Norm	3.95	1.58	0.55	3.89	1.44	0.44
Bias (%)	**Empirical**	**0.0%**	**0.2%**	**1.2%**	**0.4%**	**0.9%**	**0.8%**
	Beta	**20.3%**	**42.4%**	**85.1%**	**21.1%**	**52.6%**	**125.0%**
	Weibull	-1.5%	**0.6%**	**5.2%**	**3.3%**	**15.1%**	**40.2%**
	Mix2Norm	-3.7%	-8.0%	-11.6%	-2.5%	-6.1%	-7.3%
RMSE (%)	**Empirical**	60.1%	90.0%	142.9%	58.8%	90.4%	148.9%
	Beta	77.6%	133.6%	248.4%	73.8%	139.1%	300.3%
	Weibull	50.5%	65.5%	86.2%	50.7%	72.6%	114.9%
	Mix2Norm	62.1%	89.3%	126.9%	60.2%	87.8%	130.1%
n=30							
Average	**Empirical**	4.11	1.72	0.63	3.99	1.54	0.48
	Beta	4.45	1.98	0.80	4.35	1.91	0.76
	Weibull	4.06	1.72	0.64	4.14	1.77	0.66
	Mix2Norm	4.04	1.64	0.57	3.93	1.47	0.45
Bias (%)	**Empirical**	**0.4%**	**0.5%**	**0.7%**	**0.1%**	-0.2%	-0.5%
	Beta	**8.5%**	**15.7%**	**28.9%**	**8.9%**	**24.4%**	**58.6%**
	Weibull	-1.0%	**0.5%**	**3.6%**	**3.8%**	**14.9%**	**38.2%**
	Mix2Norm	-1.4%	-4.6%	-7.9%	-1.4%	-4.8%	-5.3%
RMSE (%)	**Empirical**	48.7%	72.7%	114.1%	46.9%	71.8%	119.6%
	Beta	51.9%	79.4%	127.6%	49.7%	84.7%	162.8%
	Weibull	40.7%	52.3%	67.2%	41.1%	58.8%	92.8%
	Mix2Norm	49.7%	72.2%	103.4%	47.8%	70.3%	105.4%

* The highlighted values represent the distributional form with the lowest absolute bias or efficiency value by district and coverage level.

\+ The bolded values represent bias values for which the fitted rate is greater than the empirical/burn rate.

Table 23: Out-of-Sample Rate Simulation Analysis: μ=160; σ=40

Data Generating Process:		***Weibull***			***beta***		
		Coverage Level[*+]			**Coverage Level**		
Statistic	**Distribution**	85%	75%	65%	85%	75%	65%
	True Theoretical	7.13	3.69	1.71	7.15	3.61	1.58
n=10							
Average	**Empirical**	7.14	3.71	1.72	7.17	3.64	1.61
	Beta	8.39	5.34	3.30	8.49	5.41	3.35
	Weibull	6.88	3.62	1.75	6.88	3.61	1.73
	Mix2Norm	6.63	3.28	1.45	6.76	3.29	1.40
Bias (%)	**Empirical**	**0.1%**	**0.4%**	**0.9%**	**0.3%**	**0.8%**	**2.0%**
	Beta	**17.7%**	**44.6%**	**92.9%**	**18.7%**	**50.0%**	**111.6%**
	Weibull	-3.5%	-1.9%	**2.0%**	-3.8%	**0.1%**	**9.3%**
	Mix2Norm	-7.0%	-11.3%	-15.4%	-5.4%	-8.7%	-11.7%
RMSE (%)	**Empirical**	70.0%	94.8%	134.2%	68.7%	93.2%	132.9%
	Beta	91.1%	142.6%	242.5%	92.4%	149.4%	268.5%
	Weibull	61.2%	75.7%	95.6%	60.5%	75.6%	99.3%
	Mix2Norm	72.6%	93.8%	121.7%	71.6%	93.1%	121.7%
n=15							
Average	**Empirical**	7.16	3.73	1.74	7.16	3.62	1.59
	Beta	8.39	5.04	2.89	8.39	5.05	2.90
	Weibull	6.95	3.64	1.74	6.98	3.65	1.74
	Mix2Norm	6.92	3.48	1.56	6.97	3.43	1.46
Bias (%)	**Empirical**	**0.4%**	**0.9%**	**1.5%**	**0.2%**	**0.3%**	**0.4%**
	Beta	**17.6%**	**36.4%**	**69.0%**	**17.3%**	**40.0%**	**83.3%**
	Weibull	-2.6%	-1.4%	**1.5%**	-2.4%	**1.3%**	**9.8%**
	Mix2Norm	-3.0%	-5.7%	-8.9%	-2.5%	-5.0%	-7.6%
RMSE (%)	**Empirical**	58.5%	78.7%	110.6%	56.6%	76.7%	109.0%
	Beta	71.8%	110.0%	178.8%	68.6%	109.2%	190.1%
	Weibull	51.2%	63.0%	78.4%	49.7%	62.5%	82.3%
	Mix2Norm	60.3%	79.0%	103.2%	58.2%	77.1%	102.6%
n=20							
Average	**Empirical**	7.11	3.68	1.70	7.15	3.58	1.55
	Beta	7.90	4.49	2.38	7.83	4.46	2.38
	Weibull	7.00	3.66	1.73	7.00	3.65	1.72
	Mix2Norm	6.97	3.52	1.57	7.04	3.45	1.47
Bias (%)	**Empirical**	-0.3%	-0.3%	-0.8%	**0.0%**	-0.7%	-1.8%
	Beta	**10.8%**	**21.5%**	**39.4%**	**9.4%**	**23.7%**	**50.9%**
	Weibull	-1.9%	-1.1%	**1.0%**	-2.1%	**1.3%**	**9.0%**
	Mix2Norm	-2.2%	-4.8%	-8.0%	-1.5%	-4.3%	-7.3%
RMSE (%)	**Empirical**	49.9%	67.3%	94.2%	47.7%	64.4%	92.1%
	Beta	53.8%	79.0%	122.1%	54.0%	83.0%	139.6%
	Weibull	43.8%	53.8%	66.5%	42.3%	52.7%	68.7%
	Mix2Norm	50.7%	67.1%	88.4%	48.6%	64.4%	86.1%
n=30							
Average	**Empirical**	7.13	3.71	1.72	7.14	3.59	1.57
	Beta	7.57	4.09	2.02	7.56	4.10	2.03
	Weibull	7.04	3.67	1.72	7.04	3.67	1.72
	Mix2Norm	7.05	3.60	1.63	7.08	3.49	1.49
Bias (%)	**Empirical**	**0.0%**	**0.5%**	**0.7%**	-0.2%	-0.5%	-0.8%
	Beta	**6.1%**	**10.7%**	**18.2%**	**5.8%**	**13.7%**	**28.7%**
	Weibull	-1.3%	-0.8%	**0.7%**	-1.6%	**1.6%**	**8.7%**
	Mix2Norm	-1.1%	-2.7%	-4.7%	-1.0%	-3.2%	-5.5%
RMSE (%)	**Empirical**	40.6%	55.2%	77.6%	39.4%	53.4%	75.8%
	Beta	42.4%	59.2%	85.6%	40.8%	59.4%	92.9%
	Weibull	35.8%	44.0%	54.2%	34.8%	43.4%	56.3%
	Mix2Norm	41.0%	54.8%	73.3%	39.6%	53.2%	71.7%

[*] The highlighted values represent the distributional form with the lowest absolute bias or efficiency value by district and coverage level.

[+] The bolded values represent bias values for which the fitted rate is greater than the empirical/burn rate.

Table 24: Out-of-Sample Rate Simulation Analysis: μ=180; σ=20

Data Generating Process:		***Weibull***			***beta***		
		Coverage Level^{*+}			**Coverage Level**		
Statistic	**Distribution**	85%	75%	65%	85%	75%	65%
	True Theoretical	1.29	0.30	0.05	1.11	0.19	0.02
n=10							
Average	**Empirical**	1.31	0.30	0.06	1.11	0.19	0.02
	Beta	2.49	1.16	0.55	2.57	1.21	0.59
	Weibull	1.32	0.35	0.08	1.41	0.38	0.09
	Mix2Norm	1.01	0.19	0.04	0.91	0.14	0.02
Bias (%)	**Empirical**	**1.1%**	-0.5%	**0.7%**	0.0%	**2.4%**	**6.0%**
	Beta	**92.7%**	**290.4%**	**912.1%**	**130.9%**	**554.3%**	**2902.9%**
	Weibull	**2.5%**	**18.4%**	**51.2%**	**26.5%**	**104.1%**	**352.6%**
	Mix2Norm	-21.8%	-34.7%	-30.9%	-18.6%	-25.1%	**4.2%**
RMSE (%)	**Empirical**	131.8%	265.7%	589.1%	130.7%	299.2%	880.1%
	Beta	244.9%	711.1%	2436.5%	304.9%	1252.2%	7571.7%
	Weibull	96.7%	156.6%	277.1%	113.0%	263.1%	797.2%
	Mix2Norm	123.7%	200.7%	385.2%	121.7%	214.4%	525.0%
n=15							
Average	**Empirical**	1.29	0.29	0.05	1.12	0.19	0.02
	Beta	2.25	0.93	0.39	2.09	0.82	0.33
	Weibull	1.32	0.34	0.07	1.41	0.36	0.08
	Mix2Norm	1.08	0.21	0.04	0.99	0.16	0.02
Bias (%)	**Empirical**	-0.5%	-1.5%	-4.3%	**0.4%**	**1.5%**	-5.0%
	Beta	**74.0%**	**211.1%**	**618.6%**	**87.9%**	**341.8%**	**1594.7%**
	Weibull	**2.2%**	**13.3%**	**35.3%**	**26.2%**	**96.5%**	**310.5%**
	Mix2Norm	-16.3%	-28.9%	-33.9%	-11.2%	-15.7%	**1.3%**
RMSE (%)	**Empirical**	107.6%	212.7%	459.8%	107.3%	239.4%	612.2%
	Beta	194.5%	536.8%	1737.3%	207.3%	773.9%	4158.6%
	Weibull	77.9%	119.7%	195.3%	93.2%	212.2%	610.5%
	Mix2Norm	104.8%	168.1%	285.0%	104.6%	187.7%	409.2%
n=20							
Average	**Empirical**	1.28	0.29	0.05	1.11	0.18	0.02
	Beta	1.84	0.64	0.23	1.81	0.61	0.22
	Weibull	1.32	0.33	0.07	1.42	0.36	0.08
	Mix2Norm	1.14	0.22	0.04	1.01	0.16	0.02
Bias (%)	**Empirical**	-1.0%	-3.7%	-11.8%	0.0%	-3.2%	-13.5%
	Beta	**42.0%**	**114.2%**	**312.2%**	**62.6%**	**230.9%**	**1002.4%**
	Weibull	**1.8%**	**10.3%**	**27.2%**	**27.6%**	**96.6%**	**301.7%**
	Mix2Norm	-11.8%	-24.8%	-26.9%	-9.0%	-15.8%	-1.9%
RMSE (%)	**Empirical**	92.3%	179.0%	366.9%	91.6%	201.1%	531.3%
	Beta	134.7%	338.9%	1001.3%	158.8%	564.0%	2896.3%
	Weibull	67.1%	101.1%	158.2%	81.9%	187.4%	524.7%
	Mix2Norm	92.4%	151.5%	268.4%	89.9%	161.5%	351.3%
n=30							
Average	**Empirical**	1.28	0.29	0.05	1.11	0.18	0.02
	Beta	1.52	0.42	0.11	1.42	0.37	0.09
	Weibull	1.31	0.32	0.07	1.43	0.36	0.07
	Mix2Norm	1.19	0.24	0.04	1.04	0.16	0.02
Bias (%)	**Empirical**	-0.9%	-1.9%	-6.8%	-0.6%	-2.9%	**0.4%**
	Beta	**17.5%**	**42.4%**	**106.7%**	**27.7%**	**98.3%**	**358.1%**
	Weibull	**1.6%**	**7.5%**	**18.9%**	**27.9%**	**93.3%**	**281.6%**
	Mix2Norm	-8.1%	-18.8%	-23.8%	-6.8%	-12.2%	**10.3%**
RMSE (%)	**Empirical**	75.2%	149.4%	308.9%	74.0%	168.7%	473.7%
	Beta	85.7%	180.6%	446.9%	90.2%	259.5%	1050.7%
	Weibull	54.1%	79.0%	117.2%	67.8%	156.8%	430.2%
	Mix2Norm	76.9%	131.1%	227.4%	73.6%	140.1%	326.6%

[*] The highlighted values represent the distributional form with the lowest absolute bias or efficiency value by district and coverage level.

[+] The bolded values represent bias values for which the fitted rate is greater than the empirical/burn rate.

Table 25: Out-of-Sample Rate Simulation Analysis: μ=180; σ=30

Data Generating Process:		***Weibull***			***beta***		
		Coverage Level^{*+}			**Coverage Level**		
Statistic	**Distribution**	85%	75%	65%	85%	75%	65%
	True Theoretical	3.58	1.35	0.43	3.43	1.15	0.30
n=10							
Average	**Empirical**	3.59	1.34	0.43	3.43	1.14	0.29
	Beta	5.18	2.88	1.59	5.18	2.86	1.58
	Weibull	3.51	1.40	0.50	3.61	1.44	0.51
	Mix2Norm	3.13	1.06	0.32	3.05	0.93	0.24
Bias (%)	**Empirical**	**0.2%**	-0.1%	-0.1%	**0.1%**	-0.6%	-2.9%
	Beta	**44.7%**	**114.2%**	**268.4%**	**51.1%**	**149.8%**	**430.0%**
	Weibull	-2.0%	**3.7%**	**15.7%**	**5.4%**	**25.2%**	**70.8%**
	Mix2Norm	-12.6%	-21.0%	-26.3%	-11.0%	-18.5%	-20.5%
RMSE (%)	**Empirical**	92.1%	145.2%	244.9%	89.5%	146.0%	261.7%
	Beta	142.2%	287.9%	643.9%	154.9%	360.7%	1012.2%
	Weibull	76.0%	104.4%	149.8%	76.2%	117.7%	209.3%
	Mix2Norm	93.2%	131.7%	191.3%	90.7%	130.6%	195.5%
n=15							
Average	**Empirical**	3.56	1.35	0.45	3.45	1.16	0.30
	Beta	4.90	2.48	1.23	4.73	2.33	1.12
	Weibull	3.54	1.39	0.48	3.63	1.43	0.49
	Mix2Norm	3.27	1.14	0.35	3.25	1.03	0.27
Bias (%)	**Empirical**	-0.6%	**0.2%**	**3.7%**	**0.6%**	**0.8%**	-0.5%
	Beta	**36.9%**	**84.3%**	**185.8%**	**38.1%**	**103.0%**	**275.5%**
	Weibull	-1.1%	**3.0%**	**11.2%**	**6.1%**	**24.6%**	**66.2%**
	Mix2Norm	-8.6%	-15.1%	-18.1%	-5.1%	-10.1%	-10.5%
RMSE (%)	**Empirical**	75.8%	121.8%	207.7%	74.2%	119.8%	216.7%
	Beta	112.6%	220.0%	470.2%	109.4%	241.4%	633.9%
	Weibull	61.0%	81.9%	112.6%	63.6%	98.6%	174.1%
	Mix2Norm	78.2%	114.8%	169.6%	75.9%	113.7%	178.9%
n=20							
Average	**Empirical**	3.58	1.34	0.43	3.41	1.14	0.29
	Beta	4.42	2.03	0.89	4.30	1.93	0.83
	Weibull	3.56	1.38	0.47	3.67	1.43	0.49
	Mix2Norm	3.41	1.20	0.36	3.28	1.03	0.27
Bias (%)	**Empirical**	-0.1%	-0.6%	-1.0%	-0.4%	-1.0%	-2.0%
	Beta	**23.4%**	**51.1%**	**107.8%**	**25.5%**	**68.4%**	**178.2%**
	Weibull	-0.7%	**2.6%**	**9.1%**	**7.0%**	**24.9%**	**64.2%**
	Mix2Norm	-4.7%	-11.2%	-16.5%	-4.2%	-9.8%	-10.3%
RMSE (%)	**Empirical**	66.5%	104.9%	172.0%	63.5%	102.4%	185.3%
	Beta	85.7%	157.2%	313.9%	83.8%	174.4%	426.4%
	Weibull	54.1%	72.2%	98.1%	54.8%	84.7%	148.3%
	Mix2Norm	68.6%	101.5%	148.4%	64.8%	96.8%	150.8%
n=30							
Average	**Empirical**	3.60	1.35	0.43	3.42	1.14	0.29
	Beta	3.94	1.62	0.60	3.86	1.54	0.56
	Weibull	3.55	1.36	0.45	3.67	1.42	0.48
	Mix2Norm	3.52	1.26	0.38	3.35	1.07	0.28
Bias (%)	**Empirical**	**0.5%**	**0.3%**	-1.1%	-0.1%	-0.1%	-1.6%
	Beta	**9.9%**	**20.0%**	**40.2%**	**12.8%**	**34.6%**	**87.2%**
	Weibull	-1.0%	**0.9%**	**4.8%**	**7.1%**	**23.8%**	**60.2%**
	Mix2Norm	-1.8%	-6.6%	-11.5%	-2.1%	-6.7%	-6.9%
RMSE (%)	**Empirical**	53.3%	84.2%	140.6%	51.6%	83.8%	150.3%
	Beta	57.6%	94.5%	165.2%	57.5%	108.3%	238.4%
	Weibull	43.1%	56.5%	74.3%	44.9%	69.7%	122.5%
	Mix2Norm	54.6%	82.5%	122.8%	52.5%	80.5%	127.0%

* The highlighted values represent the distributional form with the lowest absolute bias or efficiency value by district and coverage level.

\+ The bolded values represent bias values for which the fitted rate is greater than the empirical/burn rate.

Table 26: Out-of-Sample Rate Simulation Analysis: μ=180; σ=40

Data Generating Process:		*Weibull*			*beta*		
		Coverage Level[*+]			**Coverage Level**		
Statistic	**Distribution**	85%	75%	65%	85%	75%	65%
	True Theoretical	6.46	3.09	1.30	6.41	2.93	1.13
n=10							
Average	**Empirical**	6.47	3.11	1.31	6.42	2.96	1.17
	Beta	8.15	5.03	3.02	8.09	4.98	2.98
	Weibull	6.26	3.07	1.37	6.29	3.08	1.37
	Mix2Norm	5.92	2.68	1.07	5.97	2.61	0.98
Bias (%)	**Empirical**	**0.2%**	**0.5%**	**0.6%**	**0.2%**	**1.1%**	**3.0%**
	Beta	**26.2%**	**62.8%**	**131.4%**	**26.2%**	**69.9%**	**163.1%**
	Weibull	-3.1%	-0.6%	**5.1%**	-1.8%	**5.2%**	**20.8%**
	Mix2Norm	-8.3%	-13.2%	-18.1%	-6.8%	-10.9%	-13.2%
RMSE (%)	**Empirical**	76.5%	106.8%	157.7%	74.6%	106.1%	161.4%
	Beta	106.2%	180.4%	329.2%	103.9%	184.1%	367.4%
	Weibull	66.3%	84.1%	109.1%	64.7%	85.4%	121.4%
	Mix2Norm	79.3%	104.6%	138.5%	77.7%	104.3%	143.4%
n=15							
Average	**Empirical**	6.45	3.07	1.30	6.38	2.91	1.12
	Beta	7.87	4.53	2.50	7.74	4.39	2.38
	Weibull	6.32	3.08	1.35	6.38	3.11	1.36
	Mix2Norm	6.15	2.81	1.14	6.18	2.71	1.02
Bias (%)	**Empirical**	-0.1%	-0.5%	-0.1%	-0.4%	-0.8%	-1.3%
	Beta	**21.8%**	**46.5%**	**92.0%**	**20.8%**	**49.8%**	**109.8%**
	Weibull	-2.1%	-0.4%	**3.5%**	-0.5%	**6.0%**	**20.1%**
	Mix2Norm	-4.7%	-9.0%	-12.8%	-3.6%	-7.6%	-10.4%
RMSE (%)	**Empirical**	61.4%	86.1%	127.1%	60.2%	84.9%	126.9%
	Beta	83.4%	136.8%	240.1%	76.0%	130.8%	250.6%
	Weibull	53.3%	67.1%	85.4%	52.8%	69.2%	96.7%
	Mix2Norm	63.7%	85.7%	115.6%	61.9%	84.2%	115.3%
n=20							
Average	**Empirical**	6.46	3.07	1.29	6.41	2.92	1.13
	Beta	7.47	4.05	2.07	7.40	3.98	2.01
	Weibull	6.36	3.09	1.34	6.40	3.11	1.35
	Mix2Norm	6.31	2.91	1.18	6.31	2.80	1.06
Bias (%)	**Empirical**	**0.0%**	-0.8%	-1.4%	0.0%	-0.3%	-0.2%
	Beta	**15.6%**	**31.0%**	**58.3%**	**15.5%**	**35.8%**	**77.0%**
	Weibull	-1.5%	-0.1%	**3.0%**	-0.1%	**6.0%**	**18.9%**
	Mix2Norm	-2.3%	-5.7%	-9.6%	-1.6%	-4.3%	-6.4%
RMSE (%)	**Empirical**	54.2%	75.6%	109.9%	51.9%	73.4%	111.3%
	Beta	66.2%	104.5%	175.9%	61.5%	102.3%	188.4%
	Weibull	46.7%	58.7%	74.2%	45.3%	59.2%	82.2%
	Mix2Norm	55.5%	75.6%	102.2%	53.0%	73.3%	102.7%
n=30							
Average	**Empirical**	6.45	3.07	1.29	6.40	2.93	1.13
	Beta	6.93	3.49	1.60	6.90	3.45	1.56
	Weibull	6.39	3.09	1.33	6.43	3.11	1.34
	Mix2Norm	6.37	2.96	1.21	6.34	2.83	1.08
Bias (%)	**Empirical**	-0.1%	-0.5%	-1.2%	0.0%	-0.1%	**0.1%**
	Beta	**7.3%**	**13.1%**	**22.8%**	**7.6%**	**17.6%**	**37.4%**
	Weibull	-1.0%	-0.1%	**2.0%**	**0.4%**	**6.1%**	**18.1%**
	Mix2Norm	-1.3%	-4.1%	-7.3%	-1.1%	-3.3%	-5.1%
RMSE (%)	**Empirical**	43.0%	60.2%	88.0%	42.3%	59.9%	90.1%
	Beta	46.1%	67.4%	102.7%	44.2%	67.5%	112.8%
	Weibull	37.5%	47.0%	59.1%	37.0%	48.3%	67.2%
	Mix2Norm	43.6%	60.1%	82.7%	42.9%	59.4%	83.6%

[*] The highlighted values represent the distributional form with the lowest absolute bias or efficiency value by district and coverage level.

[+] The bolded values represent bias values for which the fitted rate is greater than the empirical/burn rate.

REFERENCES

J. Atwood, S. Shaik, and M. Watts. Are Crop Yields Normally Distributed? A Reexamination. *American Journal of Agricultural Economics*, 85(4):888–901, 2003.

Turvey C.G. and J. Zhao. Parametric and Non-Parametric Crop Yield Distributions and their Effects on All-Risk Crop Insurance Premiums. University of Guelph, Ontario, 1999.

Chakravart, Laha, and Roy. *Handbook of Methods of Applied Statistics, Volume I.* John Wiley, 1967.

S. Chen and M. Miranda. Modeling Multivariate Crop Yield Densities with Frequent Extreme Values. Paper presented at the *American Agricultural Economics Association Annual Meeting*, Denver, Colorado, 2004.

R.H. Day. Probability Distributions of Field Crop Yields. *Journal of Farm Economics*, 47(3):713–741, 1965.

P. Gallagher. U.S. Corn Yield Capacity and Probability: Estimation and Forecasting with Nonsymmetric Disturbances. *North Central Journal of Agricultural Economics*, 8(1):109–122, 1986.

O. Garcia. Simplied Method-of-Moments Estimation for the Weibull Distribution. *New Zealand Journal of Forestry Science*, 11(3):304–306, 1981.

B.K. Goodwin and A. Ker. Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts. *American Journal of Agricultural Economics*, 80(1):139–153, 1998.

D.A. Hennessy, B. Babcock, and D. Hayes. Budgetary and Producer Welfare Effects of Revenue Insurance. *American Journal of Agricultural Economics*, 79(3):1024–1034, 1997.

R. Hogg, J. McKean, and A. Craig. *Introduction to Mathematical Statistics.* Pearson Prentice Hall, 2005.

R.E. Just and Q. Weninger. Are Crop Yields Normally Distributed? *American Journal of Agricultural Economics*, 81(2):287–304, 1999.

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

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

C.H. Nelson. The Influence of Distributional Assumptions on the Calculation of Crop Insurance Premia. *North Central Journal of Agricultural Economics*, 12(1):71–78, 1990.

C.H. Nelson and P. Preckel. The Conditional Beta Distribution as a Stochastic Production Function. *American Journal of Agricultural Economics*, 71(2):370–378, 1989.

B. Norwood, M. Roberts, and J. Lusk. Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions. *American Journal of Agricultural Economics*, 86(4):1032–1043, 2004.

A.Z. Pichon. Modeling Farm-Level Yield Distributions. Master's thesis, University of Illinois at Urbana-Champaign, 2002.

R.D. Pope and R. Ziemer. Stochastic Efficiency, Normality, and Sampling Errors in Agricultural Risk Analysis. *American Journal of Agricultural Economics*, 66(1):31–40, 1984.

O.A. Ramirez. Estimation and Use of a Multivariate Parametric Model for Simulating Heteroskedastic, Correlated, Nonnormal Random Variables: The Case of Corn Belt Corn, Soybean, and Wheat Yields. *American Journal of Agricultural Economics*, 79(1):191–205, 1997.

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

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

B. Silverman. *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, 1986.

G.W. Snedecor and W.G. Cochran. *Statistical Methods, Eighth Edition.* Iowa State University Press, 1989.

M.A. Stephens. EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*, 69:730–737, 1974.

H. Wang and H. Zhang. Model-Based Clustering for Cross-Sectional Time Series Data. *Journal of Agricultural, Biological, and Environmental Statistics*, 7(1):107–127, 2002.

J.M. Wooldridge. *Introductory Econometrics.* Thomson South-Western, 2003.

F.D. Zanini. *Estimating Corn and Soybean Farm-Yield Distributions in Illinois.* PhD thesis, University of Illinois at Urbana-Champaign, 2001.