Risk in Agriculture - A Study of Crop Yield Distributions and Crop Insurance

By

Narsi Reddy Gayam

Post Graduate Diploma in Rural Management (1993)
Institute of Rural Management, Anand, India

Submitted to the Engineering Systems Division in partial fulfillment of the requirements for the degree of

Masters of Engineering in Logistics

At the

Massachusetts Institute of Technology

June 2006

Signature of the Author..
Engineering Systems Division
May 12, 2006

Certified by..
Dr.David L Brock
Director, MIT Data Center
Thesis Supervisor

Accepted by ....
Dr. Yossi Sheffi
Professor, Engineering Systems Division
Professor of Civil and Environmental Engineering
Director, MIT Center for Transportation and Logistics

ABSTRACT

Agriculture is a business fraught with risk. Crop production depends on climatic, geographical, biological, political, and economic factors, which introduce risks that are quantifiable given the appropriate mathematical and statistical methodologies. Accurate information about the nature of historical crop yields is an important modeling input that helps farmers, agribusinesses, and governmental bodies in managing risk and establishing the proper policies for such things as crop insurance. Explicitly or implicitly, nearly all farm decisions relate in some way to the expectation of crop yield.

Historically, crop yields are assumed to be normally distributed for a statistical population and for a sample within a crop year. This thesis examines the assumption of normality of crop yields using data collected from India involving sugarcane and soybeans. The null hypothesis (crop yields are normally distributed) was tested using the Lilliefors method combined with intensive qualitative analysis of the data. Results show that in all cases considered in this thesis, crop yields are not normally distributed.

This result has important implications for managing risk involving sugarcane and soybeans grown in India. The last section of this thesis examines the impact of crop yield non normality on various insurance programs, which typically assume that all crop yields are normally distributed and that the probability of crop failure can be calculated given available data.

ACKNOWLEDGEMENTS

I would like to thank my advisor and thesis supervisor, Mr. Edmund W. Schuster for the constant guidance, motivation, and insights offered during the period of this research. I am also grateful to Dr. David L. Brock, Director and Founder of MIT data center for providing me an opportunity to learn about new ways that computer scientists are using to connect data, especially in the area of spatial data as applied to agriculture and the calculation of risk.

Special thanks, to the MLOG program director – Dr. Chris Caplice for his support in various ways during the course of my studies at MIT and for his efforts in developing this thesis topic within the broader field of supply chain management.

Finally, I wish to thank my classmate at MLOG, Beatrice Nnadili, for the support and help rendered at various times during my stay at MIT.

TABLE OF CONTENTS

LIST OF FIGURES

1.0 INTRODUCTION

Agriculture is the provider of food security and is important to the economy of any country. According to the Food and Agricultural Organization,"Food security exists when all people, at all times, have access to sufficient, safe and nutritious food to meet their dietary needs and food preferences for an active and healthy life." Food security is a major determinant of national security and self sufficiency in food production and is vital for any country.

In India, agriculture and allied sectors employ about 60 % of the total work force and account for 25 % of the total GDP of the country. Since 1950, continuous improvements have occurred in irrigation, technology, application of modern agricultural practices, and availability of agricultural credit. This has been called the "green revolution" and has resulted in development of high yielding varieties. These improvements have led to significant increase in crop yield per unit area. However, in comparison to other countries, the average yield in India is still low and is equal to 30% to 50% of the highest average yield in the world.(Wikipedia)

In the US, agriculture has played an important role in the early years but its share of GDP and number of people employed has continuously decreased. According to the data compiled by the economic research service of the United States Department of Agriculture (USDA), in 1930, the agriculture sector employed 21.5 % of the total work force in the US and agriculture GDP accounted for 7.7 % of the total GDP of the country.

But, in the year 2000, 1.9 % of the total work force was employed in agriculture. In 2002, GDP of agriculture was just 0.7% of the total GDP of US (Wikipedia).

Changes in commodity prices is an important factor influencing the share of agriculture associated with GDP. The prices of agricultural commodities have increased over time, however the share of agriculture as part of GDP has decreased because of the higher production of goods and services in other sectors.

Agriculture as a business is unique. Crop production is dependent on many climatic, geographical, biological, political and economic factors that are mostly independent of one another. These multiple factors introduce risk. The efficient management of these risks is imperative for the successful agricultural and consistent output of food.

1.1. Concept of Risk in Agriculture:

The Economic Research Service of the United States department of agriculture describes five categories of risks in agriculture.

1. **Production risk***:* The quality and quantity of the commodities produced is affected by uncertainties associated with the biological growth of the crops. These uncertainties can be caused by weather patterns, pest and disease incidence, and usage of various inputs like seeds, fertilizers and pesticides.
2. **Price or market risk:** This risk derives from the fluctuations in prices that the producers receive for the commodities along with the prices paid by the farmer for inputs which will increase the cost of production.

3. Financial risk: This refers to the interest cost of the capital that the farmer invests in the production process. Availability of credit and fluctuations in the interest rates contribute to the risk.

4. Institutional risk: Government regulations dealing with subsidies, export and import regulations of commodities, tax laws, chemical usage and environmental regulations impose certain amount of risk on agribusiness.

5. Human or personal risk: This refers to the health risks associated with usage of agricultural inputs like chemicals, fertilizers and agricultural machinery that create certain personal risks to the farmers engaged in agricultural production.

As evident from this categorization, agricultural risk primarily arises from a probability of adverse effects like fluctuations in weather conditions, crop failures because diseases and pests, difficulties in planning of harvest operations, and factors like price volatility and unexpected changes in global and local trade policies. These adverse situations result in disruptions and difficulties for agribusiness operations. Effective risk management tools are necessary to estimate the probability of such unfavorable situations and to minimize the consequences. Accurate and reliable information about historical crop yields is, thus, vital for decisions relating to agricultural risk management.

Historical crop yield information is also important for the supply chain operations of companies engaged in industries that use agricultural produce as raw material. Livestock, food, animal feed, chemical, poultry, fertilizer, pesticide, seed, paper and many other industries use agricultural products as ingredients in their production processes. An accurate estimate of crop sizes and risk helps these companies in planning supply chain decisions like production scheduling, raw material procurement, and inventory management. Businesses such as seed, fertilizer, agrochemical, and agricultural machinery industries plan production and marketing activities based on crop production estimates.

Yield information also influences price movements in commodity markets. Reliable estimates provide stability to the markets, help in establishing an orderly market for buyers and sellers, and in establishing futures markets for agricultural commodities. Crop size and commodity production information has a direct influence on policies for international trade in grains and other commodities. Important funding agencies, such as the IMF and the World Bank use crop production data and yield to make decisions about loans and support programs involving governments in the third world. Public agencies make policies dealing with agricultural subsidies, incentive and crop support programs, crop insurance, procurement, stock management, and distribution of grains based on the production estimates of crop size.

However, it is observed that estimates of crop size, production and yield suffer from several important short comings.

- In developing countries like India, crop estimates are largely a product of subjective reporting of crop area and yield;

- In most countries, estimates are available only for major commodities like rice, wheat, maize, sugarcane, soybeans, and other commercially traded crops. Reliable estimates are difficult to obtain for fruit and vegetable crops, millets, and crops grown in relatively small acreages.

- Estimates are available only at national and state level. Regional or county level forecasts are rarely available.

- Often conflicting figures are reported. In some instances more than one national agency issues crop production forecasts.

Given this background, the primary objective of this thesis is to examine some of the assumptions used in crop size estimation. Further, the impact of these assumptions on risk management methods like crop insurance is evaluated.

The next section describes the assumptions and crops examined in this study.

2. RESEARCH QUESTION AND METHODOLOGY.

Historical yield is an important part of risk management and crop forecasting. The majority of risk management procedures used in agriculture assume that yield is normally distributed within a statistical population and normally distributed for a sample within a crop year. This thesis examines the validity of the hypothesis that crop yields are normally distributed and analyses the importance of this hypothesis on crop insurance and agricultural supply chain decisions.

2.1 A Supposition

The null hypothesis is defined as "In India, historical crop yields for Sugarcane and Soybeans are normally distributed for a statistical population and are normally distributed for a sample within a crop year."

2.2. Importance of the Null Hypothesis in Agriculture.

Based on the central limit theorem, most agricultural studies assume that crop yields are normally distributed. The assumption of normality or non-normality affects the estimation of probabilities for the severity and occurrence of yield short falls and surplus production. The assumption of normality attributes equal probabilities of high and low historical yields for a crop. It should be noted that crop yields can exhibit extreme variability caused by weather conditions, geological qualities of the soils, biological and genetic capabilities of the crop varieties grown and applications of inputs like fertilizers and pesticides. . Thus the likelihood of low yields and high yields may or may not be equal. Assuming normal distribution of yields and ignoring skew ness may result in under estimation or over estimation of the likely yields and will result in an inaccurate estimate of risk.

Understanding of the likelihood of yields is an important determinant of crop insurance premia. It is also critical for farm management decisions. Farmers use historical yield data to make product mix decisions, and also to determine the risk of crop loss. Thus, this null hypothesis is of great importance and influences many facets of agricultural supply chain ranging from farm operations to the establishment of insurance premiums.

2.3. Methodology.

For examining the null hypothesis and its importance in crop insurance, an extensive literature survey was conducted. Research papers in journals like American Journal of Agricultural Economics, Canadian Journal of Agricultural Economics, Agribusiness, and North Central Journal of Agricultural Economics were reviewed to analyze the findings of various researchers working in the general area of frequency distributions for historical crop yields. Important observations from the literature are presented in Section 3.0.

In addition to the literature survey, a quantitative analysis of actual yields was considered necessary to test the null hypothesis. This appears in Section 5.0. As part of this testing, data on historical crop yields for sugarcane and soybeans were collected from India for analysis. Data pertaining to the yields for the entire country and district wise data for Madhya Pradesh and Maharashtra states make up the data set used for testing the assumption of normality. Sugarcane and Soybeans crops were selected for data collection and analysis because of their economic importance to India's economy. A brief description of these crops is provided below.

2.3.1 Sugarcane:

According to the FAO commodity reports, India is the second biggest producer of sugarcane in the world after Brazil with annual production of 244,800,000 metric tones of Sugarcane in the year 2004. The USA is the 10th largest producer with an annual production of 27,501,310 metric tons. Sugarcane is used for producing different types of Sugar and Molasses. The products and by-products obtained from sugarcane are important raw materials in the food processing and chemical industries. The importance of Sugarcane is increasing because of its new uses in technologies like ethanol production for automobile fuel. Several countries such as India and Brazil have substantially reduced petroleum imports through the production of ethanol.

2.3.2 Soybeans.

According to the FAO commodity reports, Soybeans are an important crop in the USA, the biggest producer of soybeans in the world with an annual production of 85,740,952 metric tones in 2004. India is the 5th largest producer of soybeans in the world with an annual production of 7,000,000 metric tones in 2004. The economic importance of the Soybean is primarily in the feed and food industries. Soy meal is used in animal feeds because of its high protein content.

Derivatives or ingredients of soybeans are used in many manufactured foods. Tofu, miso, soy drinks and other food items are prepared from soybeans. Soy meal is used in animal feeds because of its high protein content.

Before discussing the results of the statistical tests for normality, the next section puts forth an intensive look at the literature associated with probability distributions of historical yield.

3.0 LITERATURE REVIEW

The Agricultural economics literature presents a large amount of research on crop yield distributions. A literature survey was conducted to gain a sense of the previous research conducted about the null hypothesis and understand the view points presented by earlier researchers. Articles in journals like the American Journal of Agricultural Economics, the Canadian Journal of Agricultural Economics, Agribusiness and the North central journal of agricultural economics were studied and an analysis of the important observations is presented below.

3.1. Research Supporting the Normality Assumption:

Just and Weninger (1999) have studied the county level data for Alfalfa, corn, grain sorghum, soybean and wheat crops provided by the Kansas state board of agriculture. Based on their experiments, they disagreed with the view point that crop yield distributions are non normal and argued that the evidence available to date is not enough to disprove normality of crop yields. They identified three important problems associated with the methodology of yield distribution analysis which might have resulted in rejection of normality by various authors. The problems specified are

1. Misspecification of non- random components of yield distributions.
2. Misreporting of statistical significance.
3. Use of aggregate time series data to represent farm level yield distributions.

The authors ascertained that these three problems are very serious and cancel the entire evidence that disproves normality. The authors criticized the common approach in testing normality, which ignores the deterministic component of yields, which is the conditional mean yield. Typically, a polynomial trend function is used to represent deterministic component and approximations of this component are used to test normality. They feel that detrending of the deterministic component has likely introduced skew ness and non-normal kurtosis in an erroneous manner. The authors opine that the ATS (Aggregate time series) crop yield data collected by governmental agencies creates a problem due to the averaging process employed. Averaging yields under-emphasizes farm specific variation while emphasizing region wide random effects. The randomness at farm level is extremely important for studying insurance programs and production under uncertainty. The farm specific variation can be caused by variety of factors like errors in management, farm specific resource constraints, and farm specific weather and pest conditions. The public yield data which is developed by stratified sampling process does not include or eliminates this farm specific information, and this leads to incorrect conclusions. So, this study suggests that using spatial data or farm specific data will help in proving that crop yield distributions are normal.

The authors suggest an alternate approach where farm level spatial data can be used for testing normality of farm specific variation and ATS data to be used for testing normality of region wide variation. They analyze detrended data and do not find any systematic evidence of skewness or non-normal kurtosis. They also feel that theoretical results of central limit theorem with appropriate specification of mean and variance functions will make normality plausible. The authors analyze the impact of yield distribution misspecification on insurance analysis and argue that normal distribution is not an unreasonable empirical distribution for studying crop insurance programs and production under uncertainty.

3.1.2. Research Supporting Non-normality of Crop Yields

Day (1965) worked on cotton, corn and oats crops in Mississippi State and has concluded that yield distributions in agricultural crops do not exhibit normality. The data used was from experiments with seven different fertilizer levels. For cotton and corn, the data was from 1921 to 1957 and for oats 1928 to 1957. The study begins by establishing an "a priori" expectation of non normality in field crop yields, and argues that both normality and log normality appear to be exceptions rather than the rule. His conclusions are that field crop yield distributions are generally non- normal and non- lognormal. He observed an interaction between the shape of the yield probability function and the levels of nitrogen input given to a crop. The degree of skew ness and kurtosis are found to be varying dependent on different crops and amount of available nutrients, especially Nitrogen levels. The degree of skew ness and kurtosis is found to be decreasing with increased levels of nitrogen of up to 45 pounds. He puts forward a view point that Mean estimates are not really a suitable measure for forecasting and prescription purpose. Instead it is recommended to use mode or median estimates of yields.

To test the hypothesis of normality, Pearson's test of skew ness augmented by a simple sign test. This has resulted in strong evidence, which supports asymmetry in field crop yields. Geary's test for Kurtosis has also shown non-normal peakedness. It is observed that all the cotton series were non normal and all the Oats and three of the corn series have exhibited departures from log normality. In his study, he found a strong evidence of positive skew ness in cotton and negative skew ness in oats. All the corn series examined have shown positive skew ness and six of the seven oats series have shown negative skew ness. The cotton series progressed from highly skewed and peaked J shaped curves to curves of a cocked hat variety, which are nearly symmetric. Where as oats series started out as nearly symmetric distributions close to a normal curve and end as extreme, negatively skewed functions. This property of oats is extra ordinarily favorable to farmers because it shows that above average yields are more probable and occur with more frequency than below average yields and this could be the reason why oats is a major field crop in spite of its relatively low average profitability compared with corn.

The variability of skew ness with nitrogen levels shows that with increased levels of nitrogen, average yields of cotton are increased and positive skew ness is reduced and the results are similar for corn. However, in oats nitrogen levels increase negative skew ness. Thus, increased nitrogen levels offer a more positive risk situation to farmers because expected or mean yields are increased.

Gallagher, (1987) analyzed the soybean yields in US using a time series data from 1941 to 1984. He studied the shapes of probability distributions which reflect variations caused by weather in US average yields. He concludes that US soybean yield distributions are skewed with an upper limit on output and a high chance of occasional low yields. This is because of a capacity constraint on the full yield potential of a crop. Capacity is defined as "the yield that would occur with efficient use of the given technology for controllable inputs and ideal weather." He suggests that the yield of a crop can never exceed the biological potential of the plant; hence there is an upper limit on the yield which is called capacity. But, the plant can produce low yields under weather conditions like early frost, extreme heat or blight. It is also possible that Individual farm distributions also may vary considerably with factors like fertilizer treatments, soils, climates, chemical applications and investment in harvesting equipment which will reduce a farmer's risk of extremely low yields.

The importance of skew ness in crop yield forecasting is specified by Gallagher by suggesting that ignoring skewed distribution will lead to underestimation of the most likely yields. He explains that the recognition of skewed yields will require a moderate upward revision in early season point and interval estimates for soybean yields. The chance of yields falling below the revised forecast interval is double the chance that yield will be above the forecast interval. The probability of occasionally low yields is asymmetrically high. The concept of soybean capacity is defined by plant biology and use of technology and yield variability increases over time. Hence, importance of weather factors should be seriously considered.

Ramirez, Misra, and Field (2003) have studied Corn and Soybean crops in Corn Belt in the United States. They studied yield distributions of these crops and conclude that they are non normal and left skewed. Ramirez modified the multivariate non-normal parametric modeling procedure to analyze the aggregate Corn Belt yields and concludes that annual average Corn Belt corn and soybean yields are non normally distributed and left skewed. The authors relax' the assumption of time trend linearity and use joint tests for non normality under unrestricted model specifications to avoid double jeopardy of normality. They have observed that normal model produces bounds that are incompatible with observed corn and soybean data. In Corn yields, normal model leaves no observations above the upper boundary of its 88% band and in case of Soybean, only one observation above the upper boundary of 78% band. They found that the non-normal model adheres better to the theoretically required numbers; where as normal model implies unrealistically high upper bounds. The normal model clearly over estimates the probability of very low yields and underestimates the probability of moderately low to average yields, overestimates the probability of average to moderately high yields, and underestimates the probability of very high yields. The non-normal model is more accurate than the normal in predicting observed yield frequencies in 12 of the 13 intervals and does not show substantial under or over estimation pattern.

Thus, this article reaffirms that Corn Belt Corn and soybean yields are non normally distributed and left skewed with a small 3.0% probability of making an error in this conclusion. In Texas plains dry land cotton yields, the normality hypothesis is rejected at the 1 % significance level thus showing that they are right skewed and non normal.

Ramirez, Misra, and Nelson, (2003), have studied the west Texas cotton basis and concluded that error term distribution and the conditional distribution of the west Texas cotton basis are not normal. They find that the number of observations found below and above the lower and upper bounds of the confidence bands are much closer to the theoretically expected numbers under non normal model. They also observe that, during the planting/growing season, the estimated conditional distributions for the west Texas cotton basis do not deviate from the normality as markedly as during the harvesting and marketing season.

Atwood, Shaik, Watts, (2003) conducted experiments on Kansas farm level yields. They have reexamined the concept of normality in crop yields and feel that Just and Weninger's failure to reject normality might have resulted from individually detrending and GLS adjusting each farm's yield data. They conducted normality tests for Kansas farm level yields and conclude that individually estimating trends with short-term panel data tends to bias the analysis by failing to reject normality when the underlying distribution is actually non normal. In their studies, they found that normality is generally rejected when both Just and Weninger's and ECID procedures were applied to a larger Kansas panel data set than used by Just and Weninger. The importance of this study is to the insurance industry because individual detrending and assuming normality would have substantially reduced relative insurance premiums for a large number of potential insurees in an existing insurance product. They found that assuming normality suppressed premium rates for large number of insurees.

Norwood, Roberts, Lusk, (2004) have ranked crop yield models using the procedure of out-of -sample likelihood functions and evaluated crop yield models by determining how well they describe the distribution of out of sample yields. This approach characterizes the ability of models to describe the entire distribution of yields and not just the mean. It is observed that models assuming normality were consistently out performed by competing models. However, they feel that normality should not be rejected. They suggest that more consideration should be given to the formulation for mean yield and yield variance as suggested by Just and Weninger. It is observed that for forecasting purposes, a homoskedastic normal model forecasts with better accuracy.

They studied the semi parametric model developed by Goodwin and Ker which portrays percent deviations of yield from its mean with a non parametric kernel smoother. When various models are compared, it was observed that this semi parametric model ranked highest for forecasting purposes.

Dorfman (1992) argues that a large amount of agricultural economic data is inconsistent with the assumption of normality of crop yield distributions.

3.2. Relationship of Yield Distributions and Crop Insurance.

An important purpose of understanding yield distributions is to use these expressions of probability in designing and developing crop insurance products and to make decisions related to risk of loss. The insurance companies decide the premium rates based on the assumption that crop yields are normally distributed. A literature survey is conducted to understand the relevance of this assumption.

3.2.1. Review of Literature on Crop Insurance.

Nelson (1990) suggests that premia of crop insurance products are sensitive to the assumptions used in the calculations about yield distributions. He compared the calculations based on normal distribution and beta distribution assumptions and found that the normal distribution overstates the probability of loss relative to the beta distribution, and causes premia to be higher. He suggests that in calculating crop insurance premia using normal distribution appealing to a Central Limit Theorem is inappropriate because crop insurance loss events are not independent. A more appropriate approach would be to use distributions with flexible representation of skew ness.

Goodwin and Ker (1998) suggest alternate methods for measuring yield risk and determine premiums for crop insurance contracts. They offer some non-parametric methods for yield risk measurement. They suggest that assuming normal distribution for modeling the distribution of average yields may not be correct and offer non-parametric density estimation techniques, which do not assume a particular functional form for data distributions. Instead, their method allows the data to select the most appropriate representation of the yield distribution.

Ker and Coble (May, 2003) are of the opinion that determining whether an underlying yield density is normal, beta, or a complex mixture of various parametric distributions is difficult. For insurance they propose a semi parametric estimator, which begins with parametric estimate and then corrects it non parametrically based on the data. By conducting simulations, they conclude that the semi parametric estimator with normal distribution is more efficient than the parametric models i.e. normal and beta and the standard non parametric kernel estimator.

3.3. Summary of Literature Reviewed

Seven out of the eight research articles reviewed suggest that the crop yields are non normally distributed and skewed. Just and Weninger, the authors who disagreed with this hypothesis also do not conclusively prove that crop yields are normally distributed. They suggest that normality cannot be rejected because of the inherent problems in the analysis methodology and argue that normality is a reasonable assumption for crop insurance and production decisions. The literature strongly questions the assumption of normality and accepts the alternative hypothesis that crop yield distributions are non normal. This hypothesis was examined by carrying out quantitative analysis on the yield data collected from India. The findings of the quantitative analysis are presented in the next section.

4. DATA COLLECTION AND ANALYSIS.

4.1 Sources of data.

The required data for testing the assumption of normality was collected from India. The data for soybean and sugarcane crop yields for the entire country was collected from a report titled "Agricultural statistics at a glance" published by the Directorate of economics and statistics, Ministry of agriculture, Government of India. Statistics pertaining to all-India area, production and yield per hectare for both the crops were the data obtained.

The sugarcane data obtained was for a 49 year period from 1949-50 to 1997-98. The soybean data obtained was for a 28 year period from 1970-71 to 1997-98. In addition to the all India data, state level data from Madhya Pradesh and Maharashtra states was collected from departments of agriculture of the respective states. These two states were identified for data collection because they are major producers of sugarcane and soybeans in India.

Madhya Pradesh stands first in Soybeans production in India and accounts for 75 % of the entire soybean crop produced in India (www.Indiamart.com). Maharashtra is the second largest producer of both Soybeans and sugarcane in India (www.indiamart.com). Data pertaining to the area under cultivation, annual production and yield per hectare was collected from every district in these two states. The data for the years 1999-00 to 2002-03 for Madhya Pradesh and the data for 10 year periods from 1960-61 to 1990-91 and for each year from 1990-91 to 1995-96 for Maharashtra was obtained.

4.2 Reliability of the Data:

The data was collected from the Directorate of Economics and Statistics, considered a reputed government organization within India. This organization prepares yield estimates by conducting crop cutting experiments (CCEs) taken up under scientifically designed General Crop Estimation Surveys (GCES). The crop cutting experiments involve identification and marking of experimental plots of a specified size and shape in a selected field on the principle of random sampling, threshing the produce and recording of the produce harvested for determining the percentage of recovery of the economic or marketable form of produce.

A total of 21,488 CCEs are conducted for estimating the crop yield of sugarcane in the year 2001-2002. The number of CCEs for soybean is not available. The GCES are done by carrying out stratified multi-stage random sampling design with Tehsil/Taluk (county) as strata, revenue villages within a stratum as first stage unit of sampling, survey numbers or fields within each selected village as sampling unit at the second stage and experimental plot of a specified shape and size as the ultimate unit of sampling. The Directorate of Economics and Statistics used scientific methodology for arriving at the estimates. Hence, for the purpose of our study, this information is considered reliable. However, as mentioned later, there appear to be discrepancies in some of the data that ultimately impact normality.

Identification of a suitable statistical technique is necessary to analyze the data and arrive at conclusions. Understanding of previous methodologies followed by other researchers and the merits and demerits of these different techniques helps in identification the appropriate methodology. Thus, a review of methodologies followed by earlier researchers was carried out and a summary of these techniques is provided below.

4.3. Previous Methodologies Used

Taylor estimated multivariate non-normal probability distributions by using the method of fitting hyperbolic tangent transformations of normal variants. Normality was tested using Pearson, Geary and Wilke - Shapiro tests. Based on the results of these tests for normality on corn, soybean and wheat yields from Macoupin County, Illinois, Taylor concluded that the yields indicate significant skew ness. Moss and Shonkwiler studied US corn yield data from 1930-90. A stochastic trend model (Kalman filter) was fitted to this data and inverse hyperbolic sine transformation of normal disturbances was used. They propounded a view that these yields are negatively skewed.

Ramirez conducted normality tests on U.S corn, soybean, and wheat yields from 1950-89. Multivariate nonnormal yield distributions were estimated by using hyperbolic sine transformation. He found non-normality for corn and soybeans. Nelson and Preckel studied farm level data from five Iowa counties for 1961-70. They assumed a conditional beta distribution and represented deterministic component of yields with fertilizer as an economic variable. They have concluded from their study that yields from all five counties are negatively skewed. Day used Egon Pearson's test for skew ness and kurtosis and Geary's test for kurtosis on the data collected from Delta branch of the Mississippi state experiment station. Data from field experiments of cotton, corn and oats was used for these tests. He concluded that Pearson's test augmented by a simple sign tests strongly proves that field crop yields are asymmetric and it is desirable to non normal distributions to describe the probability properties of field crops. Just and Weninger tested normality using the method of approximation with a flexible polynomial trend where the polynomial degree is determined by the data. They concluded that normality of crop yields cannot be rejected.

4.4. Methodology used for testing the NULL Hypothesis.

4.4.1. Analysis of All- India Data:

The all-India yield data for both Sugarcane and Soybeans was plotted as histograms to observe the distribution of yields. Annual average yields in Kilograms per Hectare are plotted on the X axis and frequency of the number of observations (Years) falling within that yield bin is plotted on the Y-axis. The distribution is represented in the form of a histogram and shape of the distribution is observed and conclusions are drawn regarding the probability distribution of yields. Kurtosis and skew ness are computed.

Lilliefors test for goodness of fit was conducted on the data to test the hypothesis that crop yields are normally distributed.

4.4.2 Analysis of the State Level Sata:

The state level data is available for 4 years in Madhya Pradesh and for 6 years in Maharashtra. A minimum 30 observations are required to achieve good results in statistical analysis and this limited data is deemed insufficient to analyze for testing normality across the years. Thus, the state level data is used to test whether yields in a specific year are normally distributed within an area. The average yields of all the districts with in a particular year are taken as observations. The annual average yields are plotted on the X-axis and frequency of the number of observations (districts) falling within that yield range is plotted on the Y-axis. The yield distribution was presented as a histogram and conclusions are drawn about the shape of the distribution. Lilliefors goodness-of-fit test was conducted to validate the normality hypothesis. Yield data of the most recent year available was used for the analysis in both the states and both the crops. In case of Madhya Pradesh, data for the year 2002-03 and in Maharashtra data for the year 1995-96 were used for the analysis.

4.4.3. Statistical Analysis

Probability distribution of yield data was presented in the form histograms. Histograms and test of normality were done for six data sets. These data sets include sugarcane and Soybean yield for the entire country, each district in Madhya Pradesh, and, each district in Maharashtra. Microsoft Excel 2003 software was used for data analysis. The Lilliefors goodness of fit test program provided on the World Wide Web by University of Baltimore was used to test the normality hypothesis. Lilliefors test is an adaptation of the Kolmogorov-Smirnov test and is named after Hubert Lilliefors, professor of statistics at George Washington University. It is used to test the null hypothesis that data come from a normally distributed population, when the null hypothesis does not specify which normal distribution, i.e. does not specify the expected value and variance. In the Lilliefors test, the Kolmogorov-Smirnov test is implemented using the sample mean and standard deviation as the mean and standard deviation of the theoretical (benchmark) population against which the observed sample is compared. The Lilliefors statistic is used in a goodness-of-fit of whether an observed sample distribution is consistent with normality. The statistic measures the maximum distance between the observed distribution and a normal distribution with the same mean and standard deviation as the sample, and assesses whether this distance is greater than might be accounted for by chance.

Lilliefors goodness -of-fit test for normality was done on all the 6 data sets and if the test statistic is greater than 0.05, the null hypothesis is rejected. For doing the Lilliefors test, the frequency distribution of each data set was obtained by categorizing the data into various bins. The bins and the frequency of yields falling in each bin were entered in the online Lilliefors test tool provided by University of Baltimore. This tool is written in Javascript and is freely available on the Internet. The output was a test statistic, a statement about evidence for or against normality, Kurtosis, and Skewness. The results of this test were used to decide about the shape of the distributions of yields.

5. RESULTS AND DISCUSSION

5.1. Sugarcane Yield Distributions.

5.1.1 Yield distribution of Sugarcane - All-India yields.

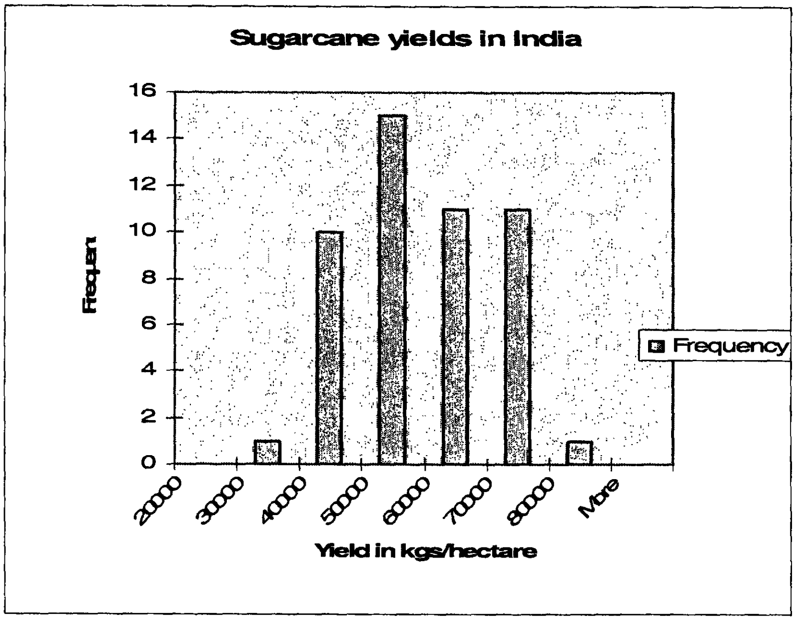

Figure 1 Distribution of sugarcane yield in India from 1949 - 1997
(Source: Directorate of Economics & Statistics, India.)

The histogram of sugarcane yields in India exhibits non-normal distribution across the years. The distribution is right skewed. The yields show variability with minimum and maximum yields of 29,495 kgs/hectare and 71,254 kgs/hectare respectively. The mean yield for the 49 year period was 50,087 kgs/hectare with a standard deviation of 11,870. The skew ness factor is -0.0228169 and kurtosis is 1.948379. The Lilliefors test for normality produced test statistic of 0.1941099. The Lilliefors test result indicates strong evidence against normality in crop yield distributions of sugarcane in India.

5.1.2. Sugarcane Yields Distributions in Madhya Pradesh in Year 2002-03.

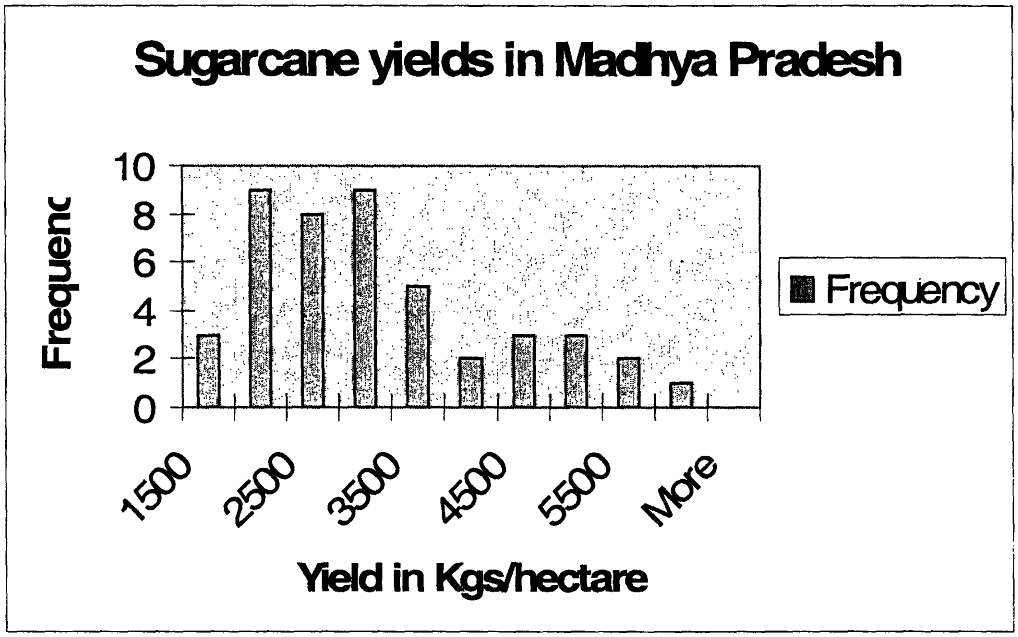

Figure 2 Sugarcane yield distribution in 45 districts of Madhya Pradesh, India (Source: Ministry of agriculture, Madhya Pradesh, India)

Sugarcane yields in 45 districts of Madhya Pradesh exhibit non-normal distribution of yields with in a specific crop year. The distribution is left skewed. The yields show high variability with a mean yield of 2,873 kgs/hectare and standard deviation of 1,141. The minimum and maximum yields are 1259 kgs / hectare and 5760 kgs/hectare respectively. Skew ness factor is 0.885 and kurtosis factor is -0.071. The Lilliefors test resulted in a test statistic of 0.1863919 and indicates strong evidence against normality.

5.1.3. Sugarcane Yield Distributions in Maharashtra in 1995-96.

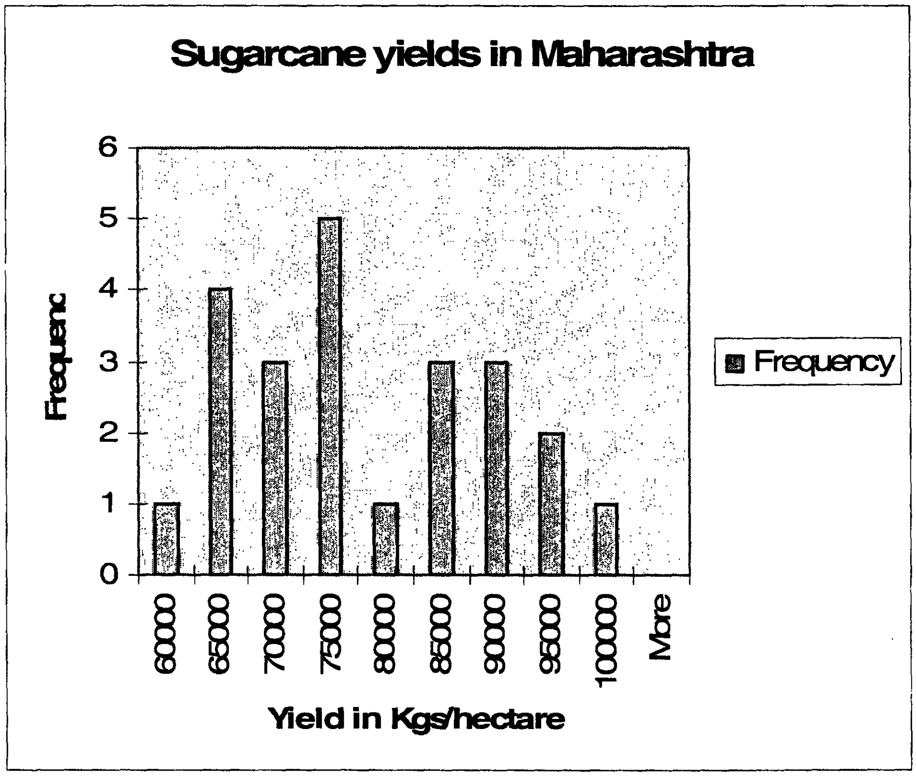

Figure 3 Sugarcane yield Distribution in 23 districts of Maharashtra (Source: Ministry of agriculture, Maharashtra, India)

The sugarcane yields in Maharashtra exhibit non-normal distribution with left skew ness. The mean yield of 23 districts for the year 1995-96 is 75,186 kgs/hectare with a standard deviation of 11,074. Maximum and minimum yields recorded are 95,896 kgs/hectare and 57,726 kgs/hectare respectively. Skew ness factor is 0.173 and kurtosis factor is -1.107. The Lilliefors test resulted in a test statistic of 0.1798393 which indicates suggestive evidence against normality.

5.2. Soybeans Yield distributions.

5.2.1. Yield Distribution of Soybean - All India yields.

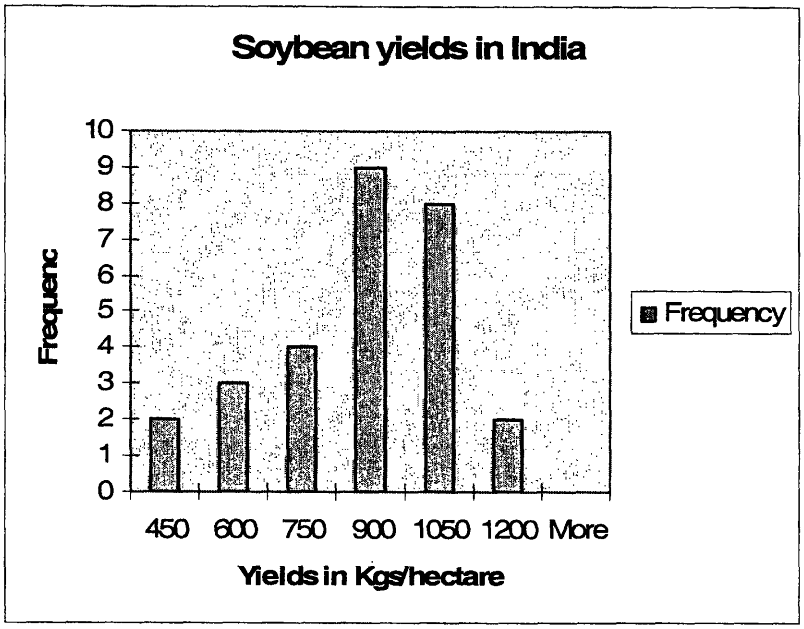

Figure 4 Soyabean yield distribution in India from 1970-1997
(Source: Directorate of Economics & Statistics, India)

The histogram for Soybean yields for 28 years in India shows that yields exhibit non-normality with right skew ness. The mean yields observed for the 28 year period display variability with a mean yield of 813 kgs/hectare and a standard deviation of 185. Kurtosis factor is -0.340 and the skew ness factor is -0.447. The maximum and minimum yields observed are 1126 kgs/hectare and 426 kgs/hectare. Lilliefors test resulted in a test statistic of 0.2213833 and indicates strong evidence against normality of crop yield distributions.

5.2.2. Soybean Yield Distributions in Madhya Pradesh in 2002-03.

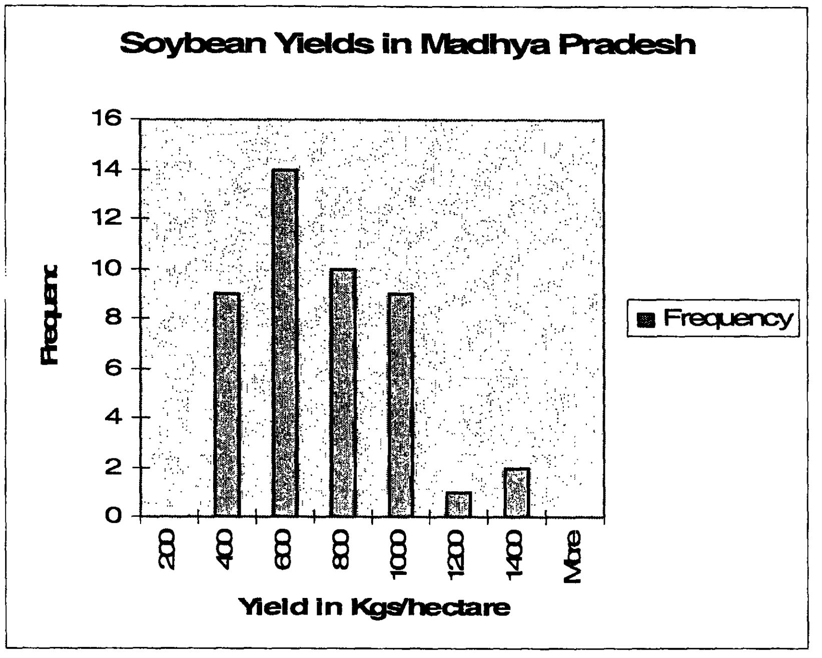

Figure 5 Soyabean yield distribution in 45 districts of Madhya Pradesh (Source: Ministry of agriculture, Madhya Pradesh)

The soybean crop yields with in a specific year in 45 districts of Madhya Pradesh exhibit a left skewed distribution. This indicates that probability of lower yields is higher than the probability of higher than mean yields. For the 45 districts of Madhya Pradesh in 2002-03 the mean was 642 kgs/hectare with a standard deviation of 253. The maximum and minimum yields recorded are 1317 kgs/hectare and 223 kgs/hectare.Skewness factor is 0.633 and Kurtosis factor is -0.088. The Lilliefors test resulted in a test statistic of 0.2034417 which indicates strong evidence against normality.

5.2.3. Soybean Yield Distribution in Maharashtra in 1995-96.

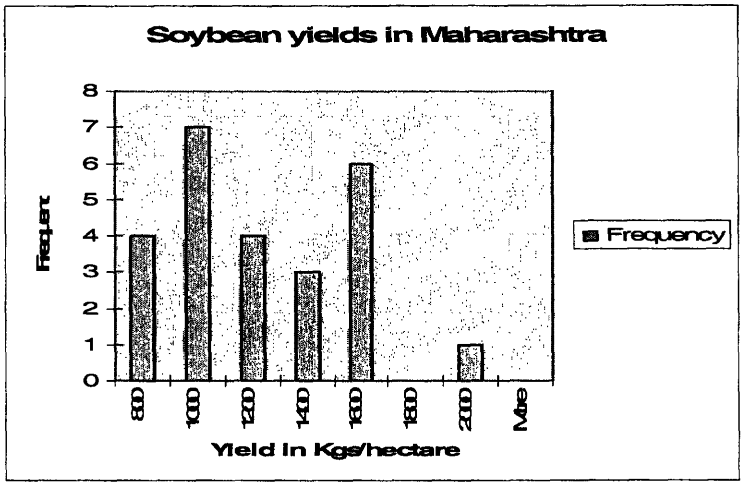

Figure 6 Soyabean yield distribution in 23 districts of Maharashtra
(Source: **Ministry of agriculture, Maharashtra)**

The analysis of the data from 25 districts of Maharashtra shows that soybean yields exhibit a skewed distribution with left skewness.This indicates that mean yields or lower than mean yields are more probable than higher than mean yields. The data shows wide variability with mean yields of 1131 kgs/hectare and a standard deviation of 324.The maximum and minimum yields recorded are 1989 kgs/hectare and 700 kgs/hectare. Skew ness factor is 0.655 and kurtosis factor is 0.356. The Lilliefors test statistic obtained was 0.2071318 which indicates strong evidence against normality of yield distributions.

5.3 Summary of the Results.

The results of the Lilliefors test and shape and appearance of yield distributions in the histograms conclusively prove that sugarcane and soybean yields in India are not normally distributed across the years. Both the crops exhibit skewed distribution with right skew ness. This indicates that the probability of higher than mean yields is more than lower than mean yields. A causal factor for skew ness could be the genetic improvements achieved in developing high yielding varieties of crops in India. The availability of high yielding seed varieties, increased use of fertilizers, improvement of irrigation facilities, improved agricultural practices due to availability of farm machinery and agricultural credit have contributed to the continuous improvements in crop yields in India. These factors have resulted in the Skewed distribution of crop yields across the years with per hectare yields increasing consistently.

However, the distribution of crop yields with-in a specific year across different locations present a different picture. Crop yields for sugarcane and soybean within a specific year in different districts of Maharashtra and Madhya Pradesh also exhibit non-normal distribution. Crop yields for both the crops in both the states exhibit strong left skew ness. This indicates that lower than mean yields are more common than higher than mean yields. The possible reasons for left skew ness of yields with in a specific year are explained below.

5.4 Reasons for Skewed Distribution of Yields.

5.4.1 Failure of the monsoon.

The yield data of Madhya Pradesh for the year 2002-03 was used for the analysis. India has experienced a severe drought continuously for a period of 3 years from 1999 to 2002. Starting from1999, India has experienced 4 consecutive years of deficient rainfall with 96 % of normal rainfall in 1999, 92% of normal rainfall in 2000 and 2001 and a scanty 81 % of normal rainfall in year 2002. This resulted in drought conditions in most of the crop growing regions. Soybean is a predominantly rain fed crop in India cultivated in small land holdings. This failure of monsoon has severely affected the soybean yields. This could be the reason for skewed distribution of yields with a higher probability of low yields.

Sugarcane is a crop cultivated under irrigation. The deficit rainfall created problems in availability of irrigation for the crop. Ground water levels were depleted and could not be replenished due to the drought. This resulted in reduction of yields in Sugarcane. This is a probable explanation for the left skewed distribution of yields in Maharashtra and Madhya Pradesh.

5.4.2. Limitations of the Data:

The following are the limitations of the data used which may have resulted in the skewed distribution of yields:

- The data available is aggregated time series data for the entire country and for the districts within each state. The production for the entire country is divided by the area under cultivation and an average yield for the country is obtained. The same procedure is followed for district level data to arrive at the average yield for each district. This yield doesn't reflect the spatial and farm level variations with in the specific growing regions such as variability in rainfall, temperature and other weather parameters, soil characteristics, crop management practices etc. Averaging the data eliminates variation and induces non-normality. Thus, availability of location specific information from individual farms is needed to provide a better understanding of the yield variability and give a reliable estimation of normality.

- The data is collected from a central government organization, which depends on state level departments for information. Authenticity of the data cannot be ascertained. There are certain discrepancies observed in the data for Maharashtra, which weakens the reliability of this information. Observation of Soybean yield data in Maharashtra indicates that exactly similar yields are reported for many districts in a crop year. In the year 1990-91, 12 of 23 districts in Maharashtra have reported a yield of 946 kgs/hectare. In 1991-92, 14 districts have reported a yield of 698 kgs/hectare. In 1992-93, 14 districts show an average yield of 990 kgs/hectare. This repetition of similar yields creates doubts about the authenticity of the data collection methodology in terms of a random sample. However, for analysis in this thesis, the data from the years 1995-96 is used as this data displays variability.

- The number of data observations available is not sufficient to conduct detailed statistical analysis. The All-India data for Soybean yields is available only for 28 years. Similarly, district level data from Maharashtra is available for only 23 districts. This data was insufficient to conduct more statistical analysis with out using t-statistics and might have resulted in erroneous results.

5.5. Observations on Yield Behavior.

Soybean yields in Madhya Pradesh were observed to be exhibiting a decreasing trend during the 4 years for which data is available. It was found that, 30 out of 45 districts in Madhya Pradesh show a gradual decline in per hectare yield of soybeans from 1999-00 to 2002-03. The yields in the other 15 districts show erratic behavior. The average yield for the entire state has decreased from 1068 kgs/hectare in 1999-00 to 652 kgs/hectare in 2002-03. This could be due to the failure of the southwest monsoon during the years 2000 to 2002. In the year 1998, India received excellent rainfall from a very good monsoon. The annual rainfall was 106 % of the normal rainfall. But, failure of monsoon in the next 4 years caused drought conditions in most of the crop growing regions and severely affected the yields and probably caused the decrease in yields.

In case of sugarcane, the average yields in Madhya Pradesh decreased from 4,378 kgs/hectare in 1999-00 to 3962 kgs/hectare in 2002-03. The yields during the 3 year period of 2000 to 2003 have remained stable at around 3,900 kgs/hectare. Sugarcane is a crop cultivated under assured irrigated conditions. The drought conditions prevailing in the state might have resulted in reduced yields but did not reduce the yields as drastically as seen in soybeans.

These observations indicate a strong correlation between crop yields and rainfall and inputs like irrigation. Availability of location specific weather data for statistical analysis would provide a better understanding of the relation between weather parameters and yields.

5.6. CONCLUSIONS:

The following conclusions are drawn, based on the analysis of data, visual observation of the histograms and results of the Lilliefors test.

- The null hypothesis is rejected in all six cases. The normality of crop yield distributions could not be proven in case of sugarcane and soybean yields in India. It is conclusively proved that crop yields for sugarcane and soybean exhibit non-normal distribution across the years for a crop and across the locations with in a specific crop year.

- Farm specific information about crop density, soil characteristics, input usage, weather data, and crop management practices will provide a more scientific analysis of this assumption and help in designing and developing crop models with higher accuracy.

- Irrigation availability and rainfall have an impact on the shape of crop yield distributions. Availability of irrigation improves the crop yields and increases the possibility of higher than mean yields and induces right skew ness. Lack of irrigation or rainfall reduces the yields and induces left skew ness.

- There is a need to collect and utilize farm specific, spatial information. This spatial data can capture locational variations better and provide better opportunities to understand the yield distributions.

- Ignoring skewed distribution of yields is not appropriate as it impacts risk management practices and decisions such as crop insurance. The influence of normality assumption on crop insurance is examined in the last section of this thesis.

6. RISK MANAGEMENT AND CROP INSURANCE.

6.1. Crop Insurance in the United States

In the US, crop insurance has emerged as an important protection to farmers against risk. Producers of specific crops can purchase insurance policies at a subsidized rate, under Federal crop insurance programs. These insurance policies make indemnity payments to producers based on current losses related to either below-average yields (crop yield insurance) or below-average revenue (revenue insurance).

Policies are sold through private insurance companies, but the USDA's Risk Management Agency (RMA) subsidizes the insurance premiums, subsidizes a portion of the companies' administrative and operating expenses, and shares underwriting gains and losses with the companies under the Standard Reinsurance Agreement. Premium subsidy rates were raised under the Agricultural Risk Protection Act of 2000, so that most farmers pay around 40 to 50 percent of the premiums. Insurance is widely available, though coverage is not available for all crops in all areas, and all types of insurance are not available for all crops. Farmers sign up for insurance prior to planting, but usually pay premiums after harvest. Several types of crop yield and revenue insurance are available. Each has some unique features which include:

6.1.1 Yield Insurance Plans

- APH (Actual Production History) coverage is the oldest and most widely available crop insurance product. It protects farmers against yield losses due to natural causes such as drought, excessive moisture, hail, wind, frost, insects, and disease. Yield coverage levels are based on a producer's expected yield, which is calculated from the farm's actual production history (average yields over the last 4 to 10 years). The farmer selects a yield coverage level, ranging from 50 to 75 percent of average yield (up to 85 percent in some areas), and an indemnity price, ranging from 55 to 100 percent of the crop price established annually by RMA. If the harvested yield is less than the insured yield (i.e., less than the yield coverage level), the farmer receives an indemnity based on the difference between the actual yield and the insured yield. The total indemnity equals this yield shortfall times the indemnity price times acres insured.

- Catastrophic (CAT) coverage provides a lower level of coverage on yield losses at a low cost to producers. It pays indemnities at a rate of 55 percent of the established price of the commodity when farm yield losses are more than 50 percent. CAT premiums are paid by RMA, but producers must pay a $100 administrative fee for each crop insured. CAT coverage is not available on all types of policies. Yield coverage above the CAT level is often referred to as "buy-up."

- Group Risk Plan (GRP) policies use county yields as the basis for determining a loss. When the county yield for the insured crop falls below the trigger level chosen by the farmer, an indemnity is paid. Yield coverage is available for up to 90 percent of the expected county yield. GRP's premiums may be lower than those for individual insurance, but an individual farmer's crop loss may not be completely covered if the county yield does not suffer a similar level of loss. This type of insurance is best suited for farmers whose crop losses typically follow the county pattern.

- Dollar Plan coverage pays for both quantity and quality yield losses and is limited to some high-value crops (e.g., fresh market tomatoes and strawberries). It guarantees a dollar amount per acre rather than a particular yield level. Both CAT and buy-up coverage are available.

6.1.2 Revenue Insurance Plans

- Crop Revenue Coverage (CRC) provides protection against gross revenue (i.e., price times yield) falling below some guaranteed level. Guaranteed revenue is equal to the farmer's elected coverage level (50 to 75 percent), times the APH yield, times the higher of (a) the base market price, which is an average of the harvest-time futures price for the month of February prior to planting; or (b) the month-long-average-harvest market price for the last month of the contract. CRC provides higher coverage in years when prices rise after planting. When a farmer's actual revenue (calculated as the actual yield times the harvest market price) is below the guaranteed revenue, CRC pays an indemnity equal to the difference between those two amounts.

- Revenue Assurance (RA) coverage is similar to CRC, with two differences. Farmers can choose between RA's "base price option," where the revenue guarantee is determined using only the preplanting price; or the "harvest price option," where the revenue guarantee may increase up to harvest time, just like CRC. The harvest price option carries a higher premium. Revenue coverage under RA is always determined using 100 percent of the base price, whereas CRC gives farmers the option of using 95 percent of the base price in exchange for a lower premium.

- Income Protection (IP) provides protection similar to RA with the base price option but requires producers to use "enterprise units." This means that the policyholder must insure all acreage for one crop in a county under a single policy (rather than having separate policies for different landlords, land sections, etc.). Premiums are lower, but IP requires that losses be across a wider area before an indemnity is paid.

- Group Risk Income Protection (GRIP) is a revenue insurance plan that uses county yields instead of farm yields when calculating revenue coverage levels and actual revenue. Farmers may select revenue coverage levels from 70 to 90 percent of expected county revenue, where county revenue is equal to the historic county yield times the relevant futures price averaged across 5 days prior to planting. Actual county revenue is calculated as the actual county yield times a month-long average of the nearby futures price at harvest time. GRIP pays indemnities only when the average county revenue for the insured crop falls below the revenue chosen by the farmer.

- Adjusted Gross Revenue (AGR) coverage insures the revenue of the entire farm rather than an individual crop by guaranteeing a percentage of average gross farm revenue, including a small amount of livestock revenue. The plan uses information from a producer's Schedule F tax forms to calculate the policy revenue guarantee. Currently, AGR is still a pilot program that is only available in selected areas

6.2. Impact of Crop Yield Distributions on Crop Insurance.

Crop insurance policies are designed on the basis of yield expectations. The insurance agencies fix premia for the policies based on the estimates of expected crop yields and occurrence of adverse conditions like hail, floods, frost or diseases and pests. Crop yield normality is assumed for estimating yield risk. Normality assumption leads to attributing equal probabilities to occurrence of low yields and high yields. This is unfavorable to farmers and insurance companies if crop yields are actually non-normal. If the yields are right skewed, it indicates that probability of high yields is more than the probability of low yields. In such case, assuming normality over estimates the probability of low yields and fixes a higher premium for the policies. This leads to farmers incurring more expense on insurance than required and may deter some farmers from buying insurance.

When crop yields are left skewed, it indicates that the probability of low yields is higher than high yields. Normality assumption underestimates the occurrence of low yields and accordingly fixes the premia. In reality, more farmers may experience lower yields and collect indemnity from the insurance companies thus causing excess financial burden to the insurance company. Thus, understanding crop yield normality or non-normality is essential for designing a crop insurance policy which can provide advantages to both the farmers and insurance companies.

APH (Actual production history) policy is one of the most widely available insurance policies in the United States. In this policy, historical yields of a farm are the criterion on that premia are decided. The farmer has to choose the level of risk coverage before he actually plants the crop. The probability of low yield is estimated based on the historical information assuming normality of yields. However, the yields may vary significantly due to factors like climatic disturbances or diseases or pest incidence. The farmers would benefit from an efficient crop forecasting system for expected yields combining climatic information with crop models. This can assist the farmers in making informed decisions about the level of coverage they can purchase and helps growers minimize the risk of yield loss and revenue loss.

6.3. Strategies for Risk Management.

- Producers can manage the risk of income loss due to climate variability by combining crop insurance with a pre-harvest marketing plan that includes strategies like hedging and forward contracting.

- About 69% of crop failures in the U.S. are because of either drought or excessive moisture (Ibarra and Hewitt, 1999). A farmer can reduce these weather and climate-based risks and can take advantage of climate forecast information to decide about insurance levels and other risk management techniques.

- Combining information about climate forecasts with crop models allow the estimation of yield potentials for the coming cropping season. Crop models that can simulate yield based on the variety that is planted, planting date, and irrigation and soil fertility (N) levels among other management practices should be developed. These models should be able to account for rainfall amounts, temperature, and solar radiation and incidence of diseases and pests. With the help of these models, it is possible to identify a planting window in which the probabilities of collecting insurance and of having a crop failure are lower. A strategy to avoid losses is to choose planting dates with low probabilities of collecting insurance, since they also represent low probabilities of yield losses. However, if planting has to occur outside of low probability windows, a higher level of coverage may be advisable.

In summary, agriculture at its base is a statistical science. With numerous risk factors, it is only through the wise application of statistics and the formulation of mathematical models that the nature of risk can be understood by farmers, agribusinesses, and public agencies. As the sophistication of data collection increases, the ability to use statistics as a practical tool in day-to-day decision-making will take on even greater importance. New computer technologies such as the M Language and Semantic Modeling will increase the possibilities for organizing and analyzing spatial data (Brock, Schuster, Allen and Kar, 2005; Brock, Schuster, and Kutz, 2006). With this new perspective, the analysis of historical crop yields will move in new directions.

BIBLIOGRAPHY AND REFERENCES

"Agricultural statistics at a glance". Directorate of economics and statistics, Government of India., August, 2004.

Atwood, J., Shaik, S., Watts, M., "Are crop yields normally distributed? A Reexamination." American journal of Agricultural Economics. 85(4) (November 2003):888-901.

Brock, D.L., Schuster, E.W., and Kutz, Sr., T.J., "An Overview of the M Language," MIT-DATACENTER-WH-OO9, January 2006.

Brock, D.L., E.W. Schuster, S.J. Allen, and P. Kar (2005), "An Introduction to Sematic Modeling for Logistical Systems," Journal of Business Logistics 26:2.

Day, R.H., "Probability distributions of field crop yields." Journal of Farm Economics, 47 (1965):713-741.

Dorfman, J.H., "Should normality be a normal assumption?" Economic Letters 42(1993):143-147.

Gallagher. P., " U.S.Soybean yields: Estimation and forecasting with nonsymmetric disturbances." American Journal of Agricultural Economics. 69(November 1987):796-803.

Goodwin, B.K., Ker, A.P., " Non parametric estimation of crop yield distributions: Implications for rating group risk crop insurance contracts." American Journal of Agricultural Economics, Vol 80, and No.1 (Feb, 1998): 139-153.

Just, R.E., Weninger, Q. "Are crop yields normally distributed." American Journal of Agricultural Economics. 81 (May 1999): 287-304.

Ker, A.P., Coble, K., "Modeling conditional yield densities." American Journal of Agricultural Economics. Vol 85 May 2003: 291-300.

Nelson, C.H., "The influence of distributional assumptions on the calculation of crop insurance premia." North Central Journal of Agricultural Economics. Vol 12, No.1 (Jan 1990): 71-78.

Norwood, B., Roberts, M.C., Lusk, J.L., "Ranking crop yield models using out-of-sample likelihood functions." American Journal of Agricultural Economics. 86(4) (November 2004):1032-1043.

Ramirez, O.A., Misra, S.K., Field, J., "Crop yield distributions revisited." American Journal of Agricultural Economics, 2003, Volume 85: 108

Ramirez, O.A., Misra, S.K., Nelson, J., "Efficient estimation of agricultural time series models with non normal dependent variables." American Journal of Agricultural Economics, 85(4) (November 2003):1029-1040.

Online resources referred

"Farm risk management: Risk in agriculture." USDA briefing room, Economic research service, US department of agriculture.

"Crop policies": Risk Management Agency US department of Agriculture.

Commodity report on Soybeans and Sugarcane. Food and Agricultural organization. www.fao.org.

www.agriwatch.com

www.indiamart.com

www.wikipedia.com

APPENDIX 1. SUGARCANE YIELD IN INDIA

YEAR	YIELDS IN KG/HECT
1953-54	31497
1954-55	36303
1955-56	32779
1956-57	33683
1957-58	34325
1958-59	37658
1959-60	36414
1960-61	45549
1961-62	42349
1962-63	40996
1963-64	46353
1964-65	46838
1965-66	43717
1966-67	40336
1967-68	40665
1968-69	49236
1969-70	49121
1970-71	48322
1971-72	47511
1972-73	50933
1973-74	51163
1974-75	49855
1975-76	50903
1976-77	53383
1977-78	56160
1978-79	49114
1979-80	49358
1980-81	57844
1981-82	58359
1982-83	56441
1983-84	55978
1984-85	57673
1985-86	59889
1986-87	60444
1987-88	60006
1988-89	60992
1989-90	65612
1990-91	65395
1991-92	66069
1992-93	63843
1993-94	67120
1994-95	71254
1995-96	67787
1996-97	66496
1997-98	69647

APPENDIX 2. SUGARCANE YIELD IN MADHYA PRADESH

DISTRICT/YEAR (YIELDS IN KGS/HECT)	1999-00	2000-01	2001-02	2002-03
JABALPUR	2437	2129	2711	2643
KATNI	2060	2060	2163	1717
BALAGHAT	4332	4331	3998	4330
CHINDWARA	6776	5082	4404	5760
SEONI	3254	2169	3252	3036
MANDLA	2120	2287	2078	2245
DINDORI	2078	1954	1954	1995
NARSINGHPUR	4277	4759	4759	4639
SAGAR	2831	2609	2608	2607
DAMOH	2649	2130	2545	2253
PANNA	1815	1492	1789	1824
TIKAMGARH	5769	6148	6055	4728
CHHATARPUR	2240	2078	2234	2237
REWA	3549	2774	3036	2525
SIDHI	2699	2747	2120	2120
SATNA	2898	2550	2600	2268
SHAHDOL	2552	2127	2552	2552
UMARIA	0	2127	2552	2127
INDORE	3173	2745	2326	2320
DHAR	1853	1852	1852	1881
JHABUA	3079	1362	2777	2777
KHARGONE	3417	2352	3137	3473
BARWANI	3753	2689	3137	3361
KHANDWA	4892	4333	4542	4542
UJJAIN	2490	1437	1788	1629
MANDSAUR	4723	4073	3778	2952
NEEMACH	4546	3778	3837	2952
RATLAM	5831	5932	5831	5102
DEWAS	1928	1988	1867	1868
SHAJAPUR	3196	3197	3151	2283
MORENA	4738	5030	4738	4082
SHEOPUR KAL	4451	4519	5110	3789
BHIND	2526	2526	2530	2524
GWALIOR	4337	3794	4391	3633
SHIVPURI	2421	2277	2458	1482
GUNA	2351	2318	2386	1259
DATIA	3474	3122	3342	2595
BHOPAL	2350	2168	2167	1987
SEHORE	4699	4337	4337	5422
RAISEN	2178	1498	1496	1495
VIDISHA	1783	1842	1839	1813
BETUL	4473	3241	4213	4343
RAJGARH	2820	2458	2567	1770
HOSHANGABAD	3396	3396	3398	3189
HARDA	3087	2674	2879	3189
AVERAGE	4378	3847	3893	3962

APPENDIX 3. SUGARCANE YIELDS IN MAHARASHTRA (YIELDS IN KGS/HECT)

DISTRICT/ YEAR	1970-71	1980-81	1990-91	1991-92	1992-93	1993-94	1994-95	1995-96
Nasik	95377	103553	111748	111430	87100	82000	92424	86728
Dhule	72387	86841	72711	70470	70430	71000	81698	74013
Jalgaon	71500	96118	80948	71530	80590	87000	85632	84780
Ahmednagar	113567	106270	90211	81240	81310	80000	90468	83590
Pune	97331	107397	89219	88930	81110	87000	89989	90632
Solapur	98916	106443	85332	69480	68400	80000	85155	85897
Satara	93236	91109	87211	82508	83340	88000	92310	95896
Sangali	68526	87874	92064	87390	82810	91000	97091	90283
Kolhapur	76401	84399	89495	93300	81300	85000	89186	80621
Aurangabad	60204	74798	79339	57490	62000	75000	80826	66447
Jalna	0	0	0	78670	67800	76000	81029	70802
Beed	69800	69545	71309	88010	71510	73000	76472	68786
Latur	0	0	74016	60640	58340	70000	63065	57726
Osmanabad	68800	73812	74020	54000	59760	70000	70356	64072
Nanded	67218	78311	74023	46600	64680	72000	77144	60065
Parbhani	58884	68608	81368	68920	76870	70000	71230	71454
Buldhana	69090	74800	77667	78890	59410	74000	78423	70650
Akola	63200	74750	77667	72410	70730	72000	72762	75609
Amravati	63500	75000	83765	67910	52810	71000	76727	85517
Yavatmal	63200	74769	74696	70200	73920	69000	73357	63400
Wardha	63000	75000	77770	72410	62000	70000	73273	74889
Nagpur	63000	75000	77750	72410	72670	70000	77143	60250
Bhandara	63333	74750	77692	72410	64070	69000	81778	67182

APPENDIX 4. SOYABEAN YIELDS IN INDIA

YEAR	YIELDS (KG/HECT)
1970-71	426
1971-72	426
1972-73	819
1973-74	829
1974-75	768
1975-76	975
1976-77	988
1977-78	940
1978-79	975
1979-80	568
1980-81	728
1981-82	741
1982-83	637
1983-84	735
1984-85	768
1985-86	764
1986-87	584
1987-88	582
1988-89	892
1989-90	801
1990-91	1015
1991-92	782
1992-93	894
1993-94	1086
1994-95	911
1995-96	1012
1996-97	987
1997-98	1126

APPENDIX 5. SOYABEAN YIELD IN MAHARASHTRA (YIELD IN KGS/HECT)

DISTRICT/YEAR	1980-81	1990-91	1991-92	1992-93	1993-94	1994-95	1995-96
Thane	348	0	0	0	0	0	0
Raigad	334	0	0	0	0	0	0
Nasik	591	946	698	990	1338	1122	1116
Dhule	541	946	698	990	1338	1433	1416
Jalgaon	531	946	698	990	1338	1830	1143
Ahmednagar	716	0	698	990	1338	886	700
Pune	569	946	698	990	1338	725	700
Solapur	374	0	698	0	1338	528	700
Satara	814	946	698	990	1084	845	1429
Sangali	585	1447	1190	1756	1590	1506	1989
Kolhapur	1021	946	1472	2050	2220	1116	1470
Aurangabad	463	0	698	990	1338	1015	1565
Jalna	0	0	0	0	1338	918	1466
Beed	266	946	698	990	1338	741	1210
Latur	0	946	0	990	1338	719	922
Osmanabad	317	0	0	0	1338	719	732
Nanded	343	0	698	990	1338	1025	1295
Parbhani	238	946	698	990	1338	1357	1155
Buldhana	329	946	698	990	1338	1196	831
Akola	299	946	698	990	1338	800	981
Amravati	434	908	526	744	1103	957	961
Yavatmal	421	946	698	990	1338	883	1457
Wardha	193	985	724	726	1361	857	981
Nagpur	433	861	543	726	1232	789	956
Bhandara	143	1070	636	819	827	743	834
Chandrapur	105	890	650	1055	1170	802	1233
Gadchiroli	0	0	650	1055	1338	806	1044

APPENDIX 6. SOYABEAN YIELD IN MADHYA PRADESH (YIELDS IN KGS/HECT)

DISTRICT/YEAR	1999-00	2000-01	2001-02	2002-03
JABALPUR	955	748	932	760
KATNI	792	495	849	460
BALAGHAT	1175	1193	1271	1214
CHINDWARA	916	824	747	754
SEONI	953	525	1109	930
MANDLA	701	365	766	578
DINDORI	949	476	767	671
NARSINGHPUR	1220	1039	1687	974
SAGAR	685	821	681	485
DAMOH	715	518	780	470
PANNA	559	418	507	384
TIKAMGARH	963	746	1146	613
CHHATARPUR	532	373	506	303
REWA	567	546	544	402
SIDHI	488	415	430	392
SATNA	618	416	669	378
SHAHDOL	543	401	488	324
UMARIA	454	363	446	385
INDORE	1486	1031	979	786
DHAR	1003	862	873	857
JHABUA	697	332	490	549
KHARGONE	416	363	518	660
BARWANI	332	375	482	579
KHANDWA	562	314	441	504
UJJAIN	1363	611	644	473
MANDSAUR	1691	583	719	558
NEEMACH	1137	648	845	473
RATLAM	1595	697	625	439
DEWAS	1147	1010	1160	986
SHAJAPUR	1213	827	620	483
MORENA	1458	1142	1460	814
SHEOPUR KAL	1350	665	1368	1317
BHIND	748	505	822	721
GWALIOR	2099	2245	2020	1039
SHIVPURI	830	821	988	223
GUNA	922	979	948	338
DATIA	668	739	1030	596
BHOPAL	956	1201	985	933
SEHORE	1086	895	1013	965
RAISEN	1125	788	784	755
VIDISHA	965	1028	933	869
BETUL	625	581	882	663
RAJGARH	998	518	606	342
HOSHANGABAD	667	986	1022	657
HARDA	1026	964	1200	847
AVERAGE	**1068**	**767**	**840**	**652**

APPENDIX 7. SAMPLE OUTPUT OF LILLIEFORS TEST

Goodness-of-fit Test: Lilliefors Test for Normality

Observation	200	400	600	800	1000	1200	1400							
Its Frequency	0	9	14	10	9	1	2							
Observation														
Its Frequency														
Observation														
Its Frequency														

CLEAR

Skewness 0.6344866 **Kurtosis** 2.9397507

Test Statistic 0.2071318

Conclusion

Strong evidence against normality = Fuerte evidenci