CROP INSURANCE PREMIUM RATE IMPACTS OF FLEXIBLE

PARAMETRIC YIELD DISTRIBUTIONS: AN EVALUATION

OF THE JOHNSON FAMILY OF DISTRIBUTIONS

by

CATHERINE P. LAWAS, B.S.

A THESIS

IN

AGRICULTURAL AND APPLIED ECONOMICS

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for
the Degree of

MASTER OF SCIENCE

Approved

Roderick Rejesus
Chairperson of the Committee

Octavio Ramirez

Thomas Knight

Accepted

John Borrelli
Dean of the Graduate School

December, 2005

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to all those people who made this thesis possible. I gratefully appreciate the guidance and insight of my thesis advisor, Dr. Roderick Rejesus, whose expertise in crop insurance and econometric modeling has benefited me a great deal in the organization of this thesis. I am grateful to the support and mentoring of my thesis co-advisors Dr. Octavio Ramirez and Dr. Thomas Knight. As respected experts in their field, their technical assistance, comments and suggestions helped me improve this thesis.

The research funding from National Research Initiative (NRI), USDA is gratefully acknowledged. Further, I would like to extend my appreciation to Dr. Bruce Sherrick of University of Illinois for giving me permission to use their farm data. The technical assistance of Ms. Tanya McDonald in GAUSS programming is gratefully appreciated as well. I also want to thank Dr. David Willis for his assistance during the absence of my supervisor.

I am thankful to all my colleagues and friends (especially Aibee, Georgia, Justin, Rebekka, Maria, Margarita, Raghu, Mauricio, Jagada, Sid, Sandeep, Jessica, Janina, Nida, Joogee, Marie, Mats and Iyel) who helped my stay in the university easier and more fun. Especially I am obliged to Iyel for providing substantive and editorial comments. The help and encouragement of Erns, Archie and Jojo to take the opportunity to attend TTU is greatly appreciated.

I would like to give my special thanks to my family (Doc, Nanay, Alex, Thelma, Rhic, Rhim, Jo, Gil, Dennis, Cheryl, Malo, Mili, Bern, all in-laws, all pamangkin and David) whose love and support enabled me to complete this work. And lastly, I am grateful to Almighty for the countless blessings.

TABLE OF CONTENTS

ABSTRACT

This study evaluated the Johnson family of distributions as a flexible parametric approach to model crop yields. Specifically, its statistical performance was compared to the most common distribution used to model yields in the literature -- the beta distribution. All distributions examined were re-parameterized such that the suitability of the candidate distributions is solely determined by the span of the skewness-kurtosis combinations allowed by a particular distribution. This re-parameterization facilitates comparison of the performance of the distributions. The parameters of each distribution were then estimated using the maximum likelihood technique. Comparison of likelihood values was used to assess the statistical performance of the distributions. Application of the procedure to a sample of Illinois farm-level corn data showed that the Johnson family of distributions seemed to be a highly flexible parametric distribution that best fits the empirical data (as compared to the beta distribution). This may be attributed to the fact that the Johnson family can theoretically account for any possible underlying mean-variance structure and a wide variety of skewness-kurtosis combinations. The economic significance of the findings was assessed by evaluating the effect of yield distribution choice on the estimation of actuarially fair insurance premiums. Results showed that the actual unsubsidized premium rates used by RMA are significantly different from the premiums estimated using the Johnson family of distributions. This is suggestive of adverse selection problems for the sample of Illinois corn farms investigated. However, when the subsidy to the current RMA premium rates were taken into consideration, the magnitude of the difference became

smaller. Hence, the subsidies implemented by the government to encourage participation of low-risk producers seem to have the positive side-effect of reducing adverse selection in the program.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AD	Anderson-Darling
APH	actual production history
ARPA	Agricultural Risk Protection Act
CDF	cumulative distribution function
FCIC	Federal Crop Insurance Corporation
KS	Kolmogorov-Smirnov
MPCI	multiple peril crop insurance
MULTIHS	hyperbolic sine distribution model
OSLLF	out-of-sample-log-likelihood function
pdf	probability density function
RMA	Risk Management Agency
SEMIPAR	semi parametric
STOCHIHS	Shonkwiler's stochastic trend model

CHAPTER I

INTRODUCTION

General Problem

The use of crop insurance as a risk management tool has grown rapidly in recent years. Since 1938, the Federal Crop Insurance Program has aimed to stabilize U.S. agricultural producers' incomes by providing a "safety-net" when low crop yields occur. Initially, the program started as an experiment and crop insurance coverage was limited to major crops in the major producing states of the U.S. The program ended the so-called "experimental" stage with the passage of the Federal Crop Insurance Act of 1980. During this period, the marketing of crop insurance policies, previously the domain of the federal government through USDA's Risk Management Agency (RMA), was expanded to include private sector participation in the delivery of the program. Hence, the U.S. crop insurance program is unique in that it is a privately-delivered but publicly supported program.

Crop insurance became a more important agricultural policy instrument when the Federal Agricultural Improvement and Reform Act of 1996 was passed. This legislation allowed the offering of several new insurance products and/or pilot projects. The program continued to grow considerably with the passage of the Agricultural Risk Protection Act of 2000 (ARPA) which mandated the expansion of crop insurance coverage to a broader range of agricultural commodities, including among others, specialty crops, nursery, livestock, forage and rangeland, and aquaculture. Today, crop insurance has become one of the principal sources of financial protection being utilized by U.S. agricultural producers. From 1997 to 2004, the program grew from insuring 182.2 to 221 million acres, a 21 percent increase over the eight year period (Insurance Information Institute, 2005). In 2006, the program is expected to provide producers with more than $42 billion in protection on approximately 220 million acres (Risk Management Agency, 2005a).

Specific Problem

In light of the expansion in availability and use of crop insurance products, one of the most important issues in this literature is with regard to what crop yield distribution to use in premium rate setting. The yield distribution represents the risk characteristics of producing a particular crop and serves as the basis for setting premium rates in crop insurance. It has been long recognized that accurate assessment of crop yield behavior is critical for setting premium rates in crop insurance. Previous studies have shown that when premiums are not accurately set, adverse selection and/or moral hazard problems could occur (Makki and Somwaru, 2002; Just et al., 1999; Coble et al.,1997). Adverse selection is caused by the failure of the insurer to accurately rate the risk of loss, while moral hazard is caused by the hidden actions of the insured which augment the risk of loss (Makki and Somwaru, 2002).

Over the years, there have been a number of yield distributions suggested for use in crop insurance premium rate setting. Typically, the distributions suggested for use are parametric in nature. For example, the early studies about this issue have suggested the use of a normal distribution (e.g. Botts and Boles, 1958). Even a recent study has espoused the use of this distribution (see Just and Weninger, 1999). But a majority of studies have suggested that a non-normal distribution may be a more appropriate characterization of crop yields. Hence, most parametric yield distributions suggested in recent years have been non-normal. Some examples of non-normal parametric distributions that have been mentioned in the past include the beta distribution (e.g., Nelson and Preckel, 1989; Tirupattur, Hauser and Chaherli, 1996; Babcock, Hart and Hayes, 2004; and Coble, et al., 1997), the gamma distribution (Gallagher, 1987), the Weibull distribution (Sherrick et al., 2004), the Burr distribution (Chen and Miranda, 2004), the log-normal distribution (Goodwin, Roberts and Coble, 2000) and the inverse hyperbolic sine (Ramirez, 1997).

However, since parametric distributions are not entirely flexible in any conceivable distributional shape, other studies have suggested the use of nonparametric and semi-parametric distributions for yield models (e.g., Ker and Goodwin, 2000; Ker and Coble, 1998; Goodwin and Ker, 1998; Featherstone and Kastens, 2000). While non-parametric and semi-parametric distributions have their own merits of flexibility, they are not particularly applicable when there are data limitations. Typically, estimation of nonparametric distributions requires large sample sizes to achieve stability of parameter estimates (Ker and Goodwin, 2000).

As a compromise to using a traditional parametric model and a fully-flexible non-parametric approach, recent work of Ramirez and McDonald (2005) proposes a highly flexible parametric distribution which can be used as an alternative to estimating non-parametric distributions when sample size is limited and/or when the researcher is more comfortable applying the standard maximum likelihood procedures for estimating a parametric distribution. They argued that the collective Johnson family of distributions -- the S_U, S_B and lognormal -- can be viewed as an extremely flexible non-normal probability density function (pdf) model accommodating a wider range of empirically possible mean-variance-skewness-kurtosis combination to approximate a wide-array of distributional shapes. The Johnson family of distribution also allows for the possibility of modeling a variety of heteroskedastic and/or autocorrelated processes, which previous parametric methods failed to consider.

While a vast number of studies have been undertaken to determine possible crop yield distributions that could be used in crop insurance, only a few studies attempted to seek empirical evidence to evaluate the performance of alternative distributional assumptions for crop yields. Norwood, Roberts and Lusk (2004) ranked the statistical performance of several parametric and semi-parametric distributions using out-of-sample log-likelihood functions. Chen and Miranda (2004) focused on comparing the statistical performance of parametric and nonparametric distributions when accounting for extreme-events such as complete crop failure. While these two studies focused on the statistical performance of the models, only the study by Sherrick, et al. (2004) investigated both the statistical performance and the premium rate impacts of using alternative yield distributions in crop insurance.

In view of the foregoing, the main objective of this study is to evaluate both the statistical performance and premium rate impacts of the Johnson family of distributions. Unlike the paper by Sherrick, et al. (2004), this study will only compare the Johnson distribution to the beta distribution, which is the most commonly used crop yield distribution in empirical work. Specifically, using farm-level corn data from Illinois, likelihood value comparisons and simulation models are used to assess the statistical performance and the premium rate impacts of the Johnson family of distributions. Results of this study extends the literature on crop yield distributions in crop insurance by providing a more flexible set of distributions that can potentially be used for premium rate setting to mitigate adverse selection in crop insurance.

CHAPTER II

LITERATURE REVIEW

Accurate premium rates are vital for an actuarially sound insurance program. Because premiums for crop and revenue insurance are designed to cover losses over time, insurers typically estimate yield and revenue distributions to show expected losses and payouts. However, premium rates can also be determined by using experience-based data such as a farm's actual loss experience and/or yields from actual production history (APH). In this study, we will limit our discussion on the determination of accurate rates that requires precise measurement of crop yield risks through proper estimation of a yield distribution. A variety of approaches to modeling and estimating crop yield distributions are available. These approaches can be categorized into two groups based on whether they use parametric or nonparametric techniques. Parametric approaches generally involve using observed yield to estimate specific parameters that describe a known probability density function, while nonparametric approaches impose minimal structure and do not rely on a known probability distribution. The first part of this section reviews the pertinent literature under the parametric approach. This is followed by a review of studies related to evaluating the performance and premium rate impacts of alternative crop yield distributions.

Parametric Crop Yield Distributions

Common parametric approaches used in crop yield analyses include the normal, beta, and gamma distributions. Use of the normal distribution in crop insurance ratemaking was first described by Botts and Boles (1958). However, from the wide range of literature on yield modeling approaches, a number of previous studies suggest that crop yield distributions are generally non-normal.

One of the non-normal distributions that gained support in the literature is the beta distribution (Day, 1965; Nelson and Preckel, 1989; Tirupattur, Hauser and Chaherli, 1996; Babcock, Hart and Hayes, 2004; and Coble, et al., 1996) as it allows a wide range of skewness and kurtosis, and has flexible representation of the response of the first three moments to changes in inputs. Day (1965) applied the beta distribution when he modeled crop yields in the Mississippi Delta region, and his results found significant positive skewness only for cotton, no significant skewness for corn, and zero or significant negative skewness for oats. The degree of skewness was found to increase with fertilizer application. He argued that the positive skewness of yields could be explained by bad weather conditions that caused a drastic reduction in yields. On the other hand, Nelson and Preckel (1989) used a conditional beta distribution and found corn yield distributions to be negatively skewed given average fertilizer use. They used farm-level corn yields in five Iowa counties and estimated the parameters of the distribution using two-stage maximum likelihood estimation. Their results showed that fertilizers have a significant impact on each of the first three moments of the distribution of corn yield. This implies that the shape of the distribution imposes some structure on the response of the moments to changes in inputs. One drawback of their approach is that standard errors or moment elasticities are difficult to obtain since elasticities are nonlinear functions of estimated parameters.

Another common choice for the non-normal distribution of crop yields is the gamma distribution. Like the beta distribution, a gamma distribution can also capture varying degrees of skewness and kurtosis, but requires relatively few parameters (Gallagher, 1987;

Kenkel, et al., 1989; and Pope and Ziemer, 1984). Gallagher (1987) estimated soybean yields using a gamma distribution and found evidence of a negatively skewed soybean yield distribution. He stated that one of the factors that contribute to the negatively skewed distributions is the declining positive and negative marginal returns to weather inputs. Thus this study confirmed Day's (1965) result that a negatively skewed temperature distribution could produce positively skewed yields.

With the growing popularity of non-normal approaches to yield distributions in the mid-1990s, a recent article by Just and Weninger (1999) attempted to renew support for the normal distribution. They questioned the common yield analyses undertaken in the past, especially with regard to the use of aggregate yield data, inflexible trend modeling, and the interpretation of the normality test results. They argued that methodological and data limitations may have lead to rejection of normality in favor of other parametric representations. In response to this study, Atwood, Shaik, and Watts (2000) used more diverse crop-region combinations in an attempt to contradict Just and Weninger's (1999) arguments for the renewed use of the normal distribution. They applied data over 200,000 producers of six crops in seven U.S. states and confirmed non-normality of crop yields.

Other parametric distributions that have been suggested to model crop yields include the Weibull, logistic, and log-normal distributions. Sherrick et al. (2004) considered the Weibull and log-normal distributions as an alternative to the normal, beta and logistic distributions. Using farm yield data of corn and soybeans for 26 farms in the Midwest, they found that the Weibull gave a negatively skewed distribution while the log-normal resulted to a positively skewed distribution.

One concern that arose from the previous literature, is that most parametric crop yield distributions (except for the Weibull) evolve as a two-parameter family, i.e. their distributions can be described by the mean and variance (Ramirez, Moss, and Boggess, 1994). This implies that these yield distributions tend to be "inflexible" because it could not exhibit any positive mean and variance in combination with any theoretically possible skewness-kurtosis combination. Another concern is that a consensus about non-normality of crop yields has not been achieved by agricultural economists. In this light, Ramirez, Moss, and Boggess (1994) proposed a multivariate non-normal parametric modeling and simulation procedure based on an inverse hyperbolic sine transformation to normality. They considered wheat, corn, and soybeans yields to estimate the distribution and found that soybean and wheat yields follow a normal distribution while corn yields do not. They highlighted several features of their model that make it attractive for modeling crop yields (as summarized by Field, 2000). First, the interdependence between random variables is an explicit part of the density function since the inverse hyperbolic transformation has an explicit interaction term. Second, it is possible to test directly for normality using the parameter estimates because the transformation is a special case of normality. Also, the advantage of known complete distribution function allows a simultaneous estimation of the interaction terms and transformation parameters. Finally, this formulation allows for the joint estimation of the trend parameters.

Ramirez (1997) extended the model in Ramirez, Moss, and Boggess (1994) by re-parameterizing the inverse hyperbolic sine transformation to develop a multivariate, non-normal density function that contains parameters that accurately and separately account for skewness, kurtosis, heteroskedasticity, and the correlation among the random variables of interest. The model was applied to corn, soybeans, and wheat in the Corn Belt. Ramirez (1997) concluded that corn and soybean yield distributions are leptokurtic and negatively skewed. Building on the Ramirez (1997) study, Ramirez, Misra and Field (2003) introduced the concept of flexible non-normal probability density function (pdf) models by expanding and refining the parameterization of Johnson S_U family densities. The model accounts for any conditional mean and variance in combination with all theoretically possible skewness-kurtosis combinations above the lognormal line. Corn Belt corn, soybean, wheat and West Texas dry land cotton yields were analyzed using the model and reaffirmed that corn and soybean yields are non-normally distributed and negatively skewed, while cotton yield distributions were positively skewed.

While the Johnson S_U family of distributions accounted only for combinations of positive or negative skewness with a positive level of kurtosis, Ramirez and McDonald (2005) further expanded the model by including the complementary S_B family of distributions. Based on Tadikamalla and Johnson (1989), they argue that the S_B family of distributions account for all theoretically possible skewness-kurtosis combinations below the lognormal line. Thus, Ramirez and McDonald (2005) argue that the collective pdf models based on the S_U, S_B, and lognormal distributions can be considered as an extremely flexible non-normal pdf model that can accommodate any possible mean-variance-skewness-kurtosis combination.

Ranking Alternative Yield Distributions and Crop Insurance Premium Rate Evaluation

Although there exist a rich body of literature in estimating crop yield distributions, less work has been done on ranking the competing yield distributions in applied research.

Ranking yield distributions will alleviate the problems of using competing distributional assumptions in applied research. To determine the "best" model, the distance between the specified model and the true distribution should be defined as the minimum of an Information Criterion (Vuong, 1989). The popular model selection criteria are Akaike Information Criterion, out-of-sample-root-mean squared error, chi-square statistic, Kolmogorov-Smirnov (KS) statistic and the Anderson-Darling (AD) statistic. Norwood, Lusk and Roberts (2002) performed simulation studies to compare the out-of-sample-log-likelihood function (OSLLF) criterion against the other popular criteria mentioned above. Their results showed that the OSLLF criterion picked the true yield distribution with a higher frequency than other methods. From this finding, Norwood, Lusk and Roberts (2003) applied OSLLF to rank the following yield distributions: gamma, beta, the Moss and Shonkwiler stochastic trend model (STOCHIHS), Ramirez's multivariate inverse hyperbolic sine distribution (MULTIHS), Goodwin and Ker's semi-parametric (SEMIPAR) model, and the normal distribution. They found that the normal distribution was consistently outperformed by the competing non-normal distributions and the semi-parametric model outperformed the others.

Due to the complexity of the OSLLF methods, other selection methods are more popular, like the AD statistic. Sherrick et al. (2004) and Chen and Miranda (2004) used the AD test, a popular distribution-free goodness-of-fit test, to determine the appropriate candidate distribution to represent yield data. Sherrick et al. (2004) defined an AD test that measures the distance between each sample point in the empirical cumulative distribution function (CDF) and the fitted probability at that point, with greater weights assigned to regions in the tails. They compared five parametric distributions, namely the normal, logistic, Weibull, beta, and log-normal distributions, and applied them to Illinois corn and soybean farm-level yield data from 1972-1999. Their findings showed that for corn yields, the beta and Weibull distributions fit either first- or second-best and the fitting performance is nearly identical between the two distributions The ranking was followed by the logistic, then the normal, and finally log-normal. For soybean data, a slight difference in results was detected. The AD test ranked logistic first with normal and log-normal as the poorest performers. Overall, based on the AD test, the log-normal and normal underperformed as compared to the Weibull, beta, and logistic distributions.

Aside from the AD test, Miranda and Chen (2004) also used a chi-square test to evaluate the performance of the Weibull, beta, and normal distributions in estimating multivariate crop yield densities with frequent extreme events like a complete crop failure. They utilized NASS county-level yield data from 1956-1997. While the AD test rejected the distributions more frequently than chi-square tests, both tests showed that the Weibull distribution fit the data best, as compared to the normal and beta distributions.

Since chi-square is not good for small samples, Turvey and Zhao (1999) employed a KS test to compare three parametric distributions (normal, gamma and beta) for a series of 13 to 19 years of data. KS statistics has small (finite) sample properties that can be considered exact with small sample size (Bradley 1968). The study simulated the models using five crops with 609 crop yields provided by Ontario Crop Insurance Commission and confirmed that the beta distribution fit best.

Among these studies only Sherrick et al. (2004) and Turvey and Zhao (1999) extended the comparison of yield models in terms of evaluating the effect of alternative crop yield distributions on the setting of crop insurance premium rates. Turvey and Zhao (1999) compared the crop insurance premiums generated from a nonparametric kernel estimate of the yield distribution, as well as the premium effects of a normal, beta, and gamma distribution. They used the minimum square root of the sum of the squared error and found that the beta and kernel distributions were consistently positively deviated, while normal and gamma premiums were consistently negatively deviated. This implies that when kernel or beta estimators are assumed premiums would tend to be biased upward, which could discourage low risk farmers from purchasing crop insurance. When normal or gamma estimators are assumed premiums would tend to be biased downward, which could encourage high risk farmers to purchase crop insurance. On the other hand, Sherrick et al. (2004) detected across farms, the average pay-out for the beta ($5.18) and Weibull ($4.65) are higher than the normal ($4.48) and logistic($3.87). Further, they found that the actual premiums are not closely correlated with the expected payouts and suggests that this may result to problems in controlling loss ratio and difficulties in controlling adverse selection. From both studies, it is apparent that the choice of distribution is important in terms of the valuation of crop insurance.

CHAPTER III

CONCEPTUAL FRAMEWORK

Agricultural production is subject to many uncertainties. Many “perils”, or causes of loss, such as adverse weather, insect infestations and plant diseases can severely reduce the yield or quality of crop and thereby affect farmer’s profit. The catastrophic nature of many crop-related perils led to the development of a variety of crop insurance plans. Standard crop yield insurance, termed “multiple peril crop insurance” (MPCI) or “actual production history” (APH) insurance pays an indemnity at a predetermined price to replace yield loss. “Group risk” yield insurance, termed the “group risk plan,” is based on the county’s yield. It pays an indemnity when the county’s average yield falls below a yield guarantee, regardless of the farmer’s actual yields. Three farm-level revenue insurance programs are available for a limited number of crops and regions: “crop revenue coverage,” “income protection” and “revenue assurance.” These contracts guarantee a minimum level of crop revenue and pay an indemnity if revenues fall below the guarantee (Ker, 2001). In this study, the discussion will be limited to yield insurance. This section focuses on the conceptual analysis of the importance of yield distributions in setting crop insurance premiums. A framework of asymmetric information in insurance markets due to inaccurate estimation of yield distributions will also be presented.

Rate Setting

Consider the simple yield insurance contract that pays indemnities if realized yields fall below some threshold that defines a guarantee. Under this contract, the crop insurance offer is composed of four basic items: the guarantee or total liability, the premium, the process by which loss is measured, and the indemnity calculation process (Risk Management Agency, 2005b). First, the guarantee and the premium must be established. Yield guarantee determines the conditions under which indemnity payments will be paid. It establishes the total liability or the maximum possible indemnity in the event of a total loss. On the other hand, the premium determines the charge for the coverage offered under the contract. The premium rate is then the total amount charged for the insurance product (the total premium) divided by the dollar amount of protection. For instance, if a farmer is required to pay $5 in premium for every $100 of crop insurance protection, the premium rate is 5%. To determine the premium rate, we need to measure the yield risk by determining the probability that a loss will occur and the expected insured loss (expected indemnity).

Following Makki and Somwaru (2001) and Goodwin and Mahul (2004), consider an insurance contract that will pay indemnities (I) when yields fall below a certain proportion λ of the expected insured yield μ

$$I = P\,(\max\{(\lambda\mu\text{-}y), 0\})$$

where P is the price at which losses are compensated, then I depends on the insurance contract (α) that is a continuous variable such that $I = I(\alpha)$. For simplicity, assume two states of nature, *no loss* and *loss*. The farmer exposed to a risk chooses a contract that maximizes expected utility:

$$U(\bullet) = (1\text{-}p)\ U(m\text{-}\pi) + p\ U(m\text{-}\ \pi\text{-}\ d + I) \qquad (3.1)$$

where $U(\bullet)$ is the von Neumann-Morgenstern utility function that is strictly increasing, strictly concave and twice differentiable function, p is the probability of incurring a loss, 1-p is the probability of not incurring a loss, m is the initial income, π is the premium and d is the loss.

To derive an equilibrium condition, assume that there are no transaction costs, the premium can be expressed as a function of α and other observable characteristics (z) such that $\pi = \pi(\alpha, z)$. Taking the first order condition of equation (3.1) will give as the optimal choice of an insurance contract

$$U'(m-\pi)/U'(m - \pi - d + I) = p\{I'(\alpha) - \pi'(\alpha,z)\} / (1-p)\pi'(\alpha,z) \qquad (3.2)$$

which implies that demand for insurance is equal to supply of insurance in equilibrium where $U'(\bullet) > 0$ is the marginal utility of income, p/(1-p) is the odds of incurring a loss, i.e. the measure of risk associated with the insurance contract, and $I'(\alpha)$ and $\pi'(\alpha,z)$ represent the indemnity and the premium at the margin, respectively.

If the price of insurance is actuarially fair, individuals would buy full insurance resulting to an equalization of incomes in the two states of nature such that,

$$U'(m - \pi) = U'(m-\pi - d + I(\alpha)). \qquad (3.3)$$

Farmers would trade income from one state of nature to another through a payment of premium (π) to the insurer in return for a guaranteed pay indemnities (I) if a loss occurs. This trading will continue until the incomes are equalized. Substituting equation (3.3) in (3.2) will yield to optimal condition for the supply for insurance:

$$(1-p)\,\pi'(\alpha,z) = p\,\{ I'(\alpha) - \pi'(\alpha,z)\} \qquad (3.4)$$

which indicates that expected benefits are equal to expected costs for the insurer. Solving equation (3.4) will give us premium rate:

$$\pi'(\alpha,z) = pI'(\alpha) \qquad (3.5)$$

which means that the fair premium is equal to the expected insured loss, E[*loss*]. Expected insured loss can be formally defined as the product of the probability that a loss will be realized times the expected loss, given the occurrence of loss

$$\begin{aligned} E[loss] &= E\max[\lambda\mu - y, 0] \\ &= \text{Prob}[y < \lambda\mu][\lambda\mu - E(y \mid y < \lambda\mu)] \\ &= \text{Prob}(loss) * (loss \mid loss\ occurs). \end{aligned} \tag{3.6}$$

This could be written as,

$$E[loss] = \int_0^{\lambda\mu} f(y)dy \left[\lambda\mu - \left(\int_0^{\lambda\mu} f(y)ydy \Big/ \int_0^{\lambda\mu} f(y)dy \right) \right] \tag{3.7}$$

$$= F(\lambda\mu)\left[\lambda\mu - \left(\int_0^{\lambda\mu} f(y)ydy / F(\lambda\mu) \right) \right]$$

where F(*y*) and *f*(*y*) are the cumulative probability distribution functions (cdf) and the probability density function (pdf) (Goodwin, 2005).

The equation above shows the importance of precise measurement of the yield density *f(y)* (and the corresponding cdf) to accurately determine the loss function and the premium rate. The premium values will differ if a different yield distribution is used. Failure to account for the right distribution of yield could result in informational asymmetry in the insurance market, affecting the actuarial soundness of crop insurance program. The effect of asymmetric information in insurance market equilibrium is discussed in the next section.

Optimal Solutions under Different Information Conditions

This section elucidates the nature of insurance market equilibria under different information conditions using the theoretical models by Rotschild and Stiglitz (1976), and supporting statements by Makki and Somwaru (2001).

First, we begin by defining equilibrium in a competitive insurance market. Equilibrium can be obtained when customers choose contracts to maximize expected utility such that (i) no contract in the equilibrium set makes negative expected profits; and (ii) there is no contract outside the equilibrium set that will make a nonnegative profit (Rotschild and Stiglitz, 1976). Based on the previous section, the optimal choice of insurance contract is where the demand for insurance is equal to the supply of insurance. We further showed that the fair premium rate can be obtained when the expected benefit of the insurer is equal to the expected cost.

Perfect Information

Figure 3.1 illustrates the optimal insurance level of a competitive insurance market with perfect information. The M_1 and M_2 axes represent income in the two states: *no loss* and *loss*. The 45° line represents the certainty line where income is equal in both states of nature. Assume that there are two individuals representing different level of risk ("low-risk" and "high-risk") who differ in their probability of loss occurrence. Let p^H and p^L be the probability of loss occurrence for high- and low-risk individual, respectively such that p^H is greater than p^L. The indifference curves UH and UL are level sets of preferences for high-risk and low-risk individuals. If insurance contracts are sold in a full information-competitive market, then the expected profits are zero. This is represented by lines EL for low risk-individual with slope $(1-p^L)/ p^L$ and EH for high-risk individual with slope $(1-p^H)/ p^H$. These lines represent the supply of insurance which is referred to as the fair-odds line. At point E, an individual starts at an initial endowment where income is equal to *m* if no loss occurs or *m -d* if a loss occurs (where *d* is the reduction in income). Individuals may reduce their exposure to the risk of loss by trading insurance contracts along the fair-odds lines. The movement of trading in lines EL and EH continues until the equilibrium is obtained where the slope of the fair-odds line is equal to the slope of the indifference curve. At equilibrium, low risk and high risk individuals are fully insured at actuarially fair rates at points *a* and *b*, respectively.

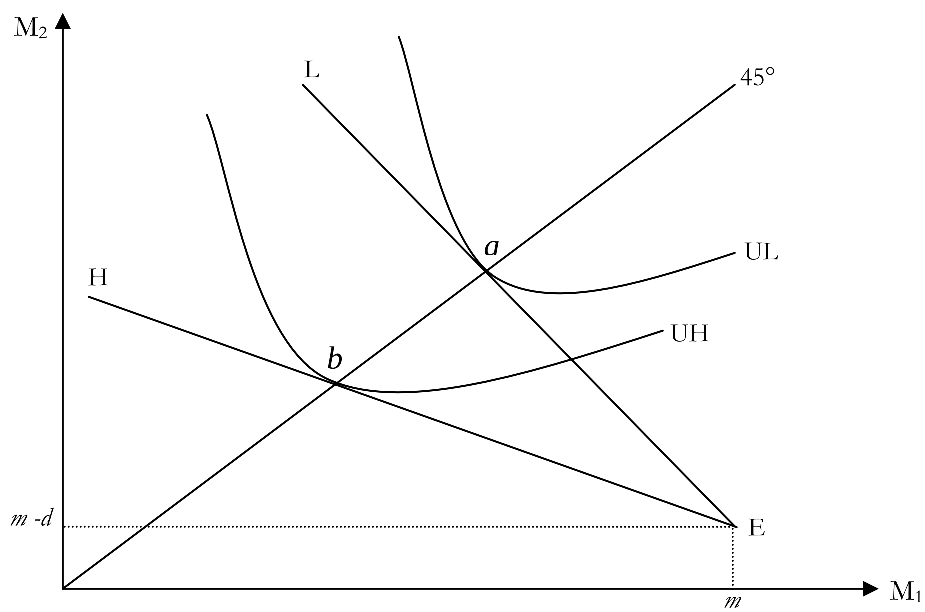


Figure 3.1. Market Equilibrium with Perfect Information

Imperfect Information

Asymmetric information has long been recognized as major cause of inefficiency in premium setting and thus the failure of crop insurance markets (Makki and Somwaru, 2002;

Just et al., 1999; Coble et al., 1997; and Chambers, 1989). Asymmetric information manifests primarily as either an adverse selection or a moral hazard problem. In the presence of adverse selection, the insured has *hidden knowledge* about the probability of loss that the insurer fails to account for in order to accurately assess the risk of loss. Therefore, the insurer is unable to set premiums commensurate with risk. Moral hazard, on the other hand is caused by *hidden actions* of the insured which increase the risk of loss and therefore the likelihood of collecting indemnities (Makki and Somwaru, 2002). For simplicity, we will consider the adverse selection problem to illustrate the existence of equilibrium in asymmetric market information. Rothschild and Stiglitz (1976) determined two kinds of equlibria in the absence of perfect information (i.e. adverse selection): *a pooling equilibrium* and a *separating equilibrium*. In a pooling equilibrium, individual risk type could not be differentiated by the insurer such that contracts are priced at an average premium and applicants buy identical contracts. On the other hand, in a separating equilibrium, different risk types can be differentiated by the insurer such that contracts are priced differently commensurate with the different risk types of the insured.

An example of a pooling equilibrium is illustrated in Figure 3.2. Based on the same market structure used in Figure 3.1 without perfect information, assume two individuals in the market: "low-risk" and "high-risk" individuals. Let EH be the contracts and UH be the indifference curves for high-risk individual. Let EL be the contracts and UL be the indifference curves for low-risk individual. When an individual's probability of loss is uncertain, *a* and *b* are not the equilibrium contract levels. Since the insurer offers an average premium (line EG), the high-risk individual will buy the contract at *b*' and the low-risk individual will purchase the contract at *a*'. At these levels of coverage, the high-risk individual is over-insured and under-priced (pays less than full insurance contracts) while the low-risk individual is under-insured and over-priced (pays more than full insurance contracts).

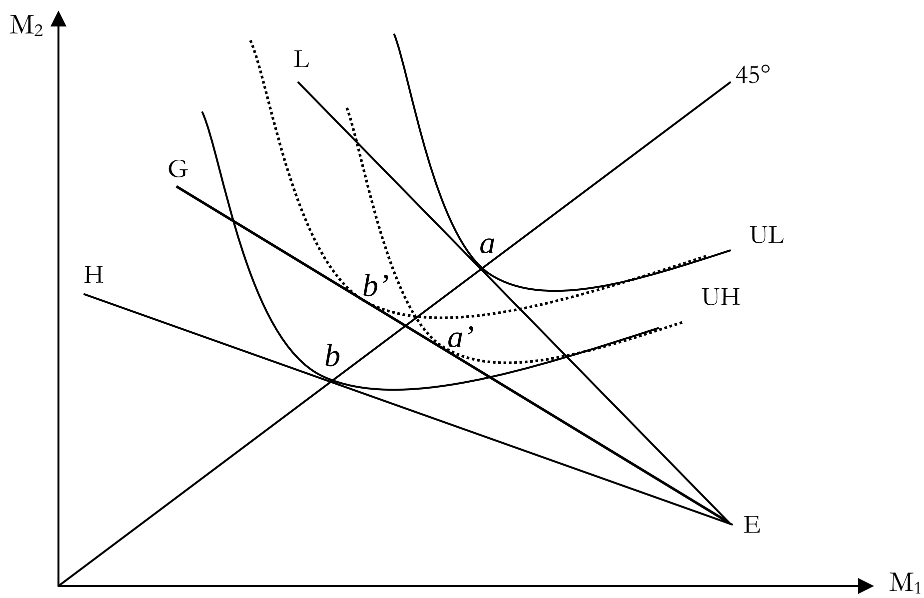


Figure 3.2. Pooling Equilibrium under Asymmetric Information

Figure 3.3 illustrates an example of a separating equilibrium. The same market structure described in Figure 3.2 will be used, except that the insurer offers two contracts at two different prices. Line EL represents the low-price contracts while line EH represents the high-price contracts. Let UL and UH be the indifference curves of low- and high-risk individuals, respectively. Similar to Figure 3.2, when individual's probability of loss is unknown, points *a* and *b* are not the equilibrium contract levels. If *a* and *b* are marketed, both high- and low-risk individuals will prefer *a* since it gives higher utility in each state. Since the insurer cannot separate low-risk individuals from high-risk individuals, the contract offered to low-risk types must be more attractive to high-risk types than their best contract. In the resulting equilibrium, high-risk types will buy contracts at a level where the high-risk fair-odds line is tangent to the high-risk indifference curves (point *b*) and the low-risk type will buy a contract at the level where low-risk fair-odds line intersects the high-risk indifference curve (point *a**). At the equilibrium, high-risk individuals buy the full insurance contract, *b* while low-risk individual will buy the partial insurance contract *a**.

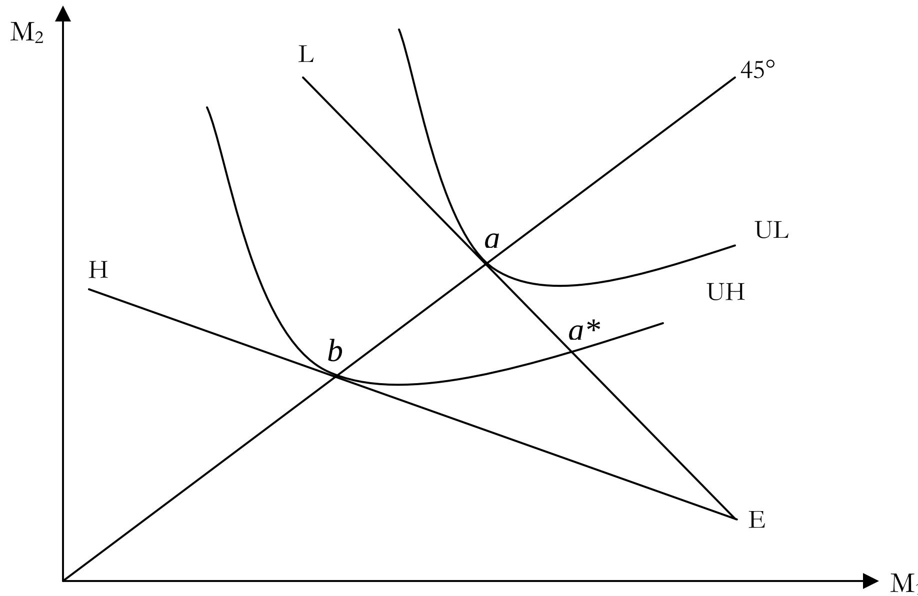


Figure 3.3. Separating Equilibrium under Asymmetric Information

In summary, when an insured's probability of loss is unknown to the insurer, the perfect information equilibrium (*a*, *b*) is not attainable. Under asymmetric information, specifically the adverse selection problem, those with high-risk individuals are over-insured and under-priced while the low-risk individuals are under-insured and over-priced.

CHAPTER IV

METHODS AND PROCEDURES

The main objective of this study is to evaluate the Johnson family of distributions (S_B, S_U and lognormal) as a flexible parametric approach for estimating crop yield distributions. To achieve this objective, we compare the statistical performance, as well as the premium rate impacts, of the Johnson family of distributions relative to the beta distribution. The beta distribution is used as the basis for comparison because it is the most popular parametric distribution used to model yields in the empirical literature (see, among others, Nelson and Preckel, 1989; Tirupattur, Hauser and Chaherli, 1996; Babcock, Hart and Hayes, 2004; and Coble, et al., 1997).

In this chapter, the procedures used to evaluate the Johnson family of distributions are discussed. The first part of this chapter discusses the different probability density functions of interest – the Johnson family and the beta distribution. In particular, the re-parameterizations of the distributions (that makes it more amenable to estimation) are described in detail. The second part of this chapter describes the estimation procedures and diagnostics needed to evaluate the candidate distributions. The third part of this chapter describes the premium rate calculations that underlie the Monte Carlo simulation model used to evaluate the premium rate impacts of using the different candidate distributions. Lastly, the final section describes the data used in the study and presents some summary statistics.

Crop Yield Models: The Johnson Family and the Beta Distribution

The Johnson family of distributions (Johnson, 1949) was proposed by Ramirez and McDonald (2005) as a flexible parametric distribution to estimate crop yield models. The system as discussed by Hoaglin, Mosteller and Tukey (1985, p.507) come about by taking a standard Gaussian random variable *Z*, translating it to $(Z-\gamma)/\delta$, and applying a transformation *T* to get

$$Y = T\left(\frac{Z-\gamma}{\delta}\right), \tag{4.1}$$

where γ and δ are distributional parameters.

The three families in the system where $x = \dfrac{Z-\gamma}{\delta}$ are (1) the family of lognormal distributions based on

$$T(x) = e^x; \tag{4.2}$$

(2) a family of bounded distributions, denoted by S_B for which

$$T(x) = \frac{e^x}{1+e^x}; \tag{4.3}$$

and (3) a family of unbounded distributions, denoted by S_U and based on

$$T(x) = \sinh(x) = \frac{1}{2}\left(e^x - e^{-x}\right). \tag{4.4}$$

In the S_B and S_U families, the parameters γ and δ determine the shape of the distribution. In lognormal family, δ governs the shape while γ influences only the scale, through the relationship $e^{(Z-\gamma)/\delta} = e^{-\gamma/\delta}e^{Z/\delta}$.

This family of distributions was employed by Ramirez (1999); Ramirez, Misra and Field (2000); Ramirez and McDonald (2005); and Ramirez and McDonald (forthcoming) to estimate non-normal crop yield distributions. Their approach was to re-parameterize the Johnson distribution so that it can accurately and separately account for skewness and kurtosis of yields, as well as model for heteroskedasticity and/or autoregressive processes. Specifically, they transformed the distribution and incorporate coefficients that would determine the mean and the variance of the distribution that is independent of its skewness and kurtosis coefficients. This procedure allows all distributional moments, including skewness and kurtosis, to arbitrarily fluctuate over time. The transformation technique used was adopted from Mood, Graybill and Boes (1974, p.200). The following are the results of the re-parameterization approach.

SU Distribution

The transformed S_U distribution is

$$z = \gamma + \delta \sinh^{-1} y \tag{4.5}$$

where $y = \frac{(x-\xi)}{\lambda}$ and the *y* must have a distribution of the same shape as *x*.

The pdf models developed by Ramirez, Misra, and Field (2000, p.109), are based on the following expanded parameterization of the S_U family of distributions:

$$Y = X\beta + \frac{\sigma\{\sinh(\Theta V) - F(\Theta, \mu)\}}{\{\Theta G(\Theta, \mu\}}, \qquad v \sim N(\mu, 1) \tag{4.6}$$

$$F(\Theta, \mu) = E[\sinh(\Theta V)] = \exp(\sinh \theta^2/2)\sinh(\Theta\mu), \quad \text{and}$$

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

where Y is the random variable of interest; X is a (1*x*k) vector of exogenous variable values shifting the mean of the Y distribution through time (t); β is a (k*x*1) vector of parameters; $\sigma^2 > 0, -\infty < \theta < \infty$ and $-\infty < \mu < \infty,$ are other distributional parameters; and sinh, cosh, and exp denote the hyperbolic sine and cosine and the exponential function, respectively. V, an independent normally distributed random variable, is the basis of the stochastic process defining the expanded S_U family of densities. Note that the last additive term of equation 4.6 represents the error term (*U*) of the model. The moments of the S_U distribution are:

$$\mathrm{E}[U] = 0, \tag{4.7}$$

$$\mathrm{Var}[U] = \sigma^2,$$

$$\mathrm{Skew}(U) = -1/4w^{1/2}(w-1)^2\left([w\{w+2\}\sinh(3\Omega)+3\sinh(\Omega)]/G(\Theta,\mu)^{3/2}\right), \text{ and}$$

$$\mathrm{Kurt}(U) = \begin{array}{l}\{1/8(w-1)^2[w^2\{w^4+2w^3+3w^2-3\}(\cosh(4\Omega)+4w^2\{w+2\}) \\ (\cosh(2\Omega)+3\{2w+1\})]\}/G(\Theta,\mu)^2\}-3\end{array},$$

where $w = \exp(\Theta^2)$ and $\Omega = -\Theta\mu$. Equation 4.7 indicates that $X\beta$ solely determines $\mathrm{E}[Y]$, σ^2 independently controls $\mathrm{Var}[U]$, and the skewness and kurtosis of the distribution of Y can be made autonomously dependent on exogenous factors by specifying μ and Θ as parametric functions of those variables. In this re-parameterization, the standard heteroskedastic specifications can be incorporated by making σ^2 as a function of the variables influencing $\mathrm{Var}[U]$, without affecting the mean or the error term of skewness or kurtosis.

This distribution can account for any conditional mean and variance in any degree of positive or negative skewness combined with positive levels of kurtosis. This means that as long as the rare negative kurtosis can be ruled out, the expanded S_U family is flexible enough to alleviate the concerns of imposing incorrect distributional assumptions when using it to estimate a true, unknown-yield distribution (Ramirez, Misra and Field, 2003).

Lognormal Distribution

The transformed lognormal distribution for a random variable x is

$$z = \gamma' + \delta \log\left(\frac{x-\xi}{\lambda}\right), \quad (4.8)$$

where $\gamma' = \gamma - \delta \log \lambda,$ and z is a unit normal variable γ.

The expanded model of the lognormal distribution developed by Ramirez and McDonald (2005) is

$$Y = X\beta + \sigma([\exp\{V-\gamma\}/\delta] - F(\delta))/\sqrt{G(\delta)}, \quad V \sim N(\mu,1) \quad (4.9)$$

$$F(\delta) = \exp(1/2\,\delta^{-2}) \text{ and}$$

$$G(\delta) = \exp(2\delta^{-2}) - \{\exp(1/2\,\delta^{-2})\}^2,$$

where variables are defined the same way as the S_U distribution such that Y is the random variable of interest, crop yield in this case; X is a vector of exogenous variables that affect the mean of the Y distribution through time (t); β is a (k$\times$1) vector of parameters; $\sigma^2 > 0$, $\delta > 0$ $\gamma = 0$ and $-\infty < \mu < \infty$, are other distributional parameters; exp denotes the exponential function; and V is an independent normally distributed random variable. The last additive term in equation 4.9 again represents the error term (*U*) of the model. The moments of this distribution are:

$$\mathrm{E}[U] = 0, \tag{4.10}$$

$$\mathrm{Var}[U] = \sigma^2,$$

$$\mathrm{Skew}(U) = \left\{\exp\left[\left(9/2\delta^{-2}\right) - 3\gamma\delta^{-1}\right] - 3\exp\left(2\delta^{-2}\right)\left(F(\delta) + 2\right)F(\delta)^3\right\} / G(\delta)^{3/2}, \text{ and}$$

$$\mathrm{Kurt}(U) = \left\{\begin{matrix}\exp\left[\left(8\delta^{-2}\right) - 4\gamma\delta^{-1}\right] - 4\left(\exp\left[\left(9/2\delta^{-2}\right) - 3\gamma\delta^{-1}\right]\right) \\ F(\delta) + 6\left(\exp\left(2\delta^{-2}\right)\right)^2 - 3\left(F(\delta)\right)^4\end{matrix}\right\} / G(\delta)^2 - 3.$$

Equation 4.10 shows that $X\beta$ solely determines E[Y], σ^2 independently controls Var[U], and δ determines skewness and kurtosis. In the same manner as the S_U distribution described previously, the standard heteroskedastic specifications can be introduced by making σ^2 as a function of the variables influencing Var[U], without affecting the mean or the error term of skewness or kurtosis. This distribution accounts in any degree of possible skewness-kurtosis combinations within the lognormal line.

<u>SB Distribution</u>

Given $y = \frac{(x-\xi)}{\lambda}$, the transformed S_B distribution is $z = \gamma + \delta \log\{y/(1-y)\}$.

Following Ramirez, Misra, and Field (2003) re-parameterization technique, the expanded S_B family distribution was developed by Ramirez and McDonald (2005). The pdf models are based on the following re-parameterized S_B family of distributions:

$$Y = X\beta + \frac{\sigma \exp\left\{\left(\frac{V-\gamma}{\delta}\right) \Big/ 1 + \exp\left(\frac{V-\gamma}{\delta}\right)\right\} - F(\Theta,\mu)}{\sqrt{G(\Theta,\mu)}}, \qquad V \sim N(\mu,1) \tag{4.11}$$

$$F(\delta,\gamma) = F1(F2 - F3)/F4 \quad \text{and}$$
$$G(\delta,\gamma) = F(\delta,\gamma) + (\delta F5) - F(\delta,\gamma)^2$$

where

$$F1 = \frac{\exp\left(-1/2\gamma^2\right)}{\sqrt{2\pi}},$$
$$F2 = 1/2\left(\delta^{-1}\right) + \delta^{-1}\sum_i \left[\left(\exp\left(i^2\right)/2\delta^2\right)\cosh\left(\left[i(\gamma\delta)\right]/\left(2\delta^2\right)\right)\left(1/\cosh\left(i/\left(2\delta^2\right)\right)\right)\right],$$
$$F3 = \sum_i (2\pi\delta)\left(\exp\left(-1/2\left((2x-1)\pi\delta\right)^2\right)\left(\sin\left((2x-1)\pi\delta\right)^2\delta\right)\right)\left(1/\left(\sinh\left((2x-1)\pi\delta\right)^2\pi\delta\right)\right),$$
$$F4 = 1 + 2\sum_i \exp\left[(-2i\pi\delta)^2\cosh(-4i\pi\delta\gamma)\right],$$
$$F5 = \left\{\left(-\gamma\left(\exp\left(-1/2\gamma^2\right)\right)/\sqrt{2\pi}\right)(F2-F3)F4 + \left((F6-F7)F1F4\right) - \left(F8F1(F2-F3)\right)\right\}/F4^2,$$
$$F6 = \sum_i \delta^{-1}\exp\left(-i^2/2\delta^2\right)\left(1/\cosh\left(i/2\delta^2\right)\right)\left(\sinh\left(i(1-2\gamma\delta)/2\delta^2\right)\right)\left(-i/\delta\right),$$
$$F7 = \sum_i (2\pi\delta)\left(\exp\left(-1/2\left((2i)-1\right)\pi\delta\right)^2\left(1/\left(\sinh\left((2i)-1\right)\pi\delta\right)\pi\delta\right)\right)\left(\cos\left((2i)-1\right)\pi\delta\gamma\right)\left((2i)-1\right)\pi\delta, \text{ and}$$
$$F8 = \sum_i 2\exp\left(-2(i\pi\delta)^2\right)\left(-\sin(i\pi\delta\gamma)2i\pi\delta\right).$$

Likewise, the variables are defined similar to S_U re-parameterization such that Y is the random variable of interest; X is a (1*x*k) vector of exogenous variable; and β is a (k*x*1) vector of parameters; $\sigma^2 > 0, -\infty < \gamma < \infty$ and $\delta > 0$ are other distributional parameters; and sinh, cosh, and exp denote the hyperbolic sine and cosine and the exponential function, respectively; and V an independent normally distributed random variable. The last additive term of equation 4.11 represents the error term (*U*) of the model. Although it is possible to express the kurtosis and skewness parameters that are independent of variance and mean parameters, we will not show the distribution's moments due to its complexity in structure (refer to Ramirez and McDonald, 2005 for more details). It is interesting to note that since the variance is independent of skewness and kurtosis parameters, heteroskedastic specification is also applicable. This model can account for any conditional mean and variance in any degree of positive or negative skewness combined with negative levels of kurtosis.

Beta Distribution

The beta distribution, being a continuous and unimodal distribution, is appropriate in crop yield modeling when a significant amount of distribution mass is expected to be located around the mean yields (Babcock, Hart and Hayes, 2004). It is one of the most popular distributions in empirical work (Day, 1965; Nelson and Preckel, 1989; Tirupattur, Hauser and Chaherli, 1996; Babcock, Hart and Hayes, 2004; and Coble, et al., 1996). Furthermore, Nelson and Preckel (2001) applied this distribution in yield estimation due to its flexibility in skewness not found in normal, log-normal, exponential and gamma distributions. They argued that the beta distribution is mathematically tractable and flexible because the first three moments responds to changes in inputs. However, Chen and Miranda (2004) argued that the beta distribution is not suited for crop yield modeling since it tends to flatten and take unreasonable "U-shapes" when the variance of the distribution rises. Such characteristics, according to Chen and Miranda (2004), are not in accordance with agronomic expectations. While the beta distribution has its own weaknesses, it is one of the most popular distributions used in yield modeling in empirical work. As such, it makes sense to compare the statistical performance and premium rate impacts of using the beta distribution against the Johnson family of distributions.

The unconditional beta probability density function of x can be written as:

$$P_x(x) = \frac{1}{B(\alpha_1,\alpha_2)} \frac{(x-\theta)^{\alpha_1-1}(\theta+\sigma-x)^{\alpha_2-1}}{\sigma^{\alpha_1+\alpha_2-1}} \qquad \theta < x < \theta+\sigma \tag{4.12}$$

$$= 0 \qquad x \geq \theta+\sigma \quad \text{or} \quad x \leq \theta$$

where θ is known lower threshold parameter, σ is scale parameter ($\sigma>0$), α_1 is the first shape parameter ($\alpha_1 > 0$) and α_2 is the second shape parameter ($\alpha_2 > 0$).

To compare this distribution to the Johnson system, a re-parameterization procedure consistent with the approach described above for the Johnson family was also employed for the beta. In doing so, we avoid the potential inconsistency in Norwood, Roberts and Lusk (2004) where they evaluated statistical performance of alternative distributions without re-parameterizing the distributions for comparability. Specifically, they compared the S_U distribution (referred as multivariate inverse hyperbolic function MULTIHS) with a linear function of time to an alternative model with a non-linear function of time. This comparison of models where the effect of time varies from one model to another may not be theoretically appropriate. This specification issue was recognized by Ramirez and McDonald (forthcoming).

Following the re-parameterization approach discussed previously, the beta pdf can then be expressed as follows:

$$Y = X\beta + \sigma\left(\left\{V - \left(\frac{\alpha_1}{\alpha_1 + \alpha_2}\right)\right\} / \sqrt{\mu - \left(\frac{\alpha_1}{\alpha_1 + \alpha_2}\right)^2}\right) \quad V \sim N(\mu, 1) \qquad (4.13)$$

where Y is the random variable of interest; X is a (1*x*k) vector of exogenous variable; and β is a (k*x*1) vector of parameters; $\sigma^2 > 0$, $\alpha_1 > 0$, and $\alpha_2 > 0$ are other distributional parameters; and V an independent normally distributed random variable. The last additive term of equation 4.13 represents the error term (*U*) of the model. The first four moments of the beta distribution are then calculated as:

$\mathrm{E}[U] = 0,$ (4.14)

$\mathrm{Var}[U] = \sigma^2,$

$$\text{Skew}(U) = \frac{\left[\frac{\Gamma(\alpha_1+3)\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(3+\alpha_1+\alpha_2)} - 3\left\{\frac{\Gamma(\alpha_1+2)\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(2+\alpha_1+\alpha_2)}\right\}\left\{\frac{\alpha_1}{\alpha_1+\alpha_2}\right\} + 2\left\{\frac{\alpha_1}{\alpha_1+\alpha_2}\right\}^3\right]}{\left\{\frac{\Gamma(\alpha_1+2)\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(2+\alpha_1+\alpha_2)}\right\}^{3/4}}, \text{ and}$$

$$\text{Kurt}(U) = \frac{\left[\frac{\Gamma(\alpha_1+4)\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(4+\alpha_1+\alpha_2)} - 4\left\{\frac{\Gamma(\alpha_1+3)\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(3+\alpha_1+\alpha_2)}\right\}\left\{\frac{\Gamma(\alpha_1+2)\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(2+\alpha_1+\alpha_2)}\right\}\left\{\frac{\alpha_1}{\alpha_1+\alpha_2}\right\}^2 - 3\left\{\frac{\alpha_1}{\alpha_1+\alpha_2}\right\}^4\right]}{-3\left\{\frac{\Gamma(\alpha_1+2)\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(2+\alpha_1+\alpha_2)}\right\}^2}.$$

The equations above indicated that $X\beta$ is determined only by E[Y], σ^2 independently controls Var[U] and α_1 and α_2 determines the skewness and kurtosis of the distribution of Y. In this specification, σ^2 can be a function of time without affecting the mean or the error term of skewness or kurtosis. This distribution has the flexibility to cover both positively and negatively skewed distribution (Nelson and Preckel, 1989).

From the framework of the re-parameterization approach, the possible ranges of skewness-kurtosis combination are presented in Figure 4.1. The family of distributions permits positive and negative skewness with associated kurtosis values resulting in a wide upper region in the plane. In particular, the lognormal is the curved red line segment in the plane. The S_U distribution covers the green plane with the combination of extreme positive kurtosis and a range of negative and positive skewness found in the middle of the plane (above lognormal area). Finally, the S_B distribution permits the negative skewness and kurtosis which is represented by blue and the yellow areas (below lognormal area). The beta distribution, on the other hand, is only represented by yellow shaded area that occupies a fairly wide plane within the S_B coverage region. Theoretically, the Johnson family of distributions can accommodate a wide plane, which suggests that it could be a highly flexible model.

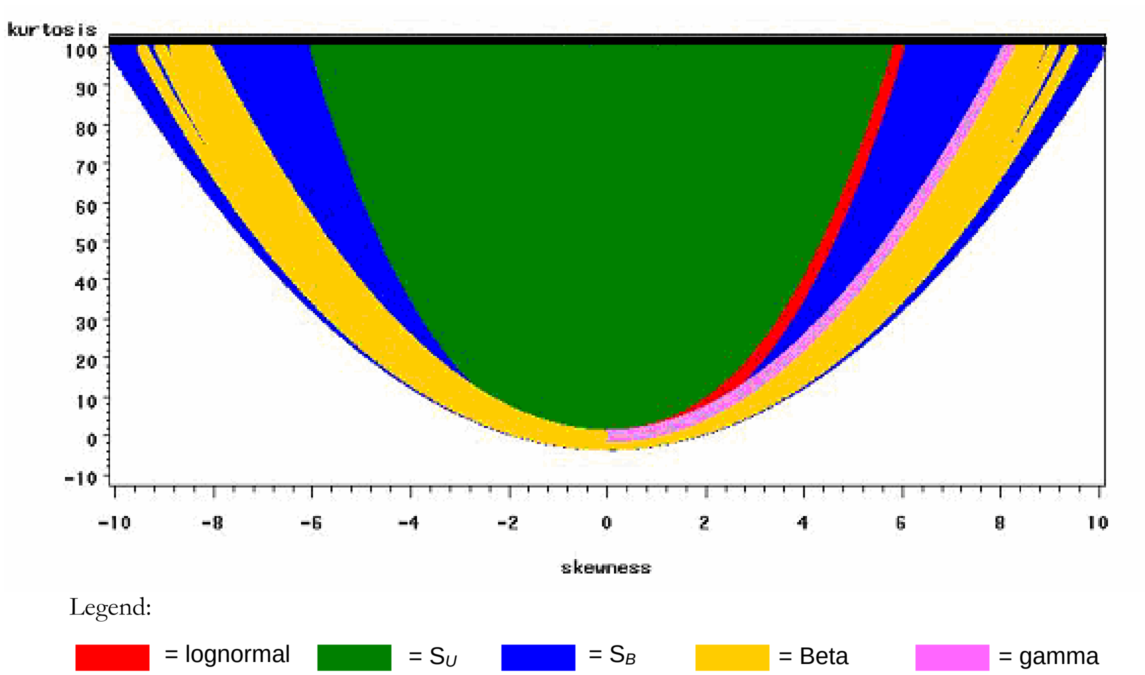


Figure 4.1. Skewness-Kurtosis Plane for Lognormal, S_U, S_B and Beta Distributions
Source: Ramirez and McDonald, forthcoming.
Note: Beta plane is a subset of the S_B plane. Gamma distribution is not included in this study.

Estimation Procedures and Diagnostics

Estimation of corn yield was determined for each farm. From the corn yield model, the mean function, was specified as a second degree polynomial function of time, t,

$$F(t) = \beta_0 + \beta_1 t + 0.01\beta_2 t^2 . \quad (4.15)$$

The coefficient of $\beta_2 t^2$ is multiplied by 0.01 to facilitate convergence in the maximum likelihood estimation procedure (described in detail below). The variance function was specified as:

$$G(t) = \sigma_0 + \sigma_1 t \,, \tag{4.16}$$

which allows for yield variances to change differently through time and address potential heteroskedasticity problems. The parameters of each of the parametric distributions were determined by maximizing the following log-likelihood function:

<u>SU:</u>

$$\mathrm{L}_{\mathrm{SU}} = -0.5\ln|\psi| + \sum_{i=1}^{n}\ln(G_i) - \sum_{i=1}^{n} 0.5 H_i^2 \tag{4.17}$$

where

$$G_i = G(\Theta,\mu)\left(1+R_i^2\right)^{-1/2}/\sigma\,,$$
$$H_i = \left\{\sinh^{-1}(R_i)/\Theta\right\} - \mu\,,$$
$$R_i = \left\{G(\Theta,\mu)\Theta U_i^* / \sigma\right\} + F(\Theta,\mu),$$

and i = 1,…, n refers to the observations, $\sinh^{-1}(x) = \ln\{x + (1+x^2)^{1/2}\}$ is the inverse hyperbolic sine function and $F(\Theta,\mu)$ is given in equation 4.6.

<u>Lognormal</u>:

$$\mathrm{L}_{\mathrm{lognormal}} = n\ln\left(\left|\delta\sqrt{G_i/2\pi}\right|\right) - \sum_{i=1}^{n}\ln\left(\left|y-\mu\sqrt{G_i}+\sigma F_i\right|\right) - 1/2\sum_{i=1}^{n}\left[\delta\ln\left(y-\ln\left(\left(\mu\sqrt{G_i}+\sigma F_i\right)/\sigma\right)\right)\right]^2 \tag{4.18}$$

where

$G_i = G(\delta)$ and $F_i = F(\delta)$ which are given in equation 4.9 and i = 1,…, n refers to the observations.

<u>SB</u>:

$$\mathrm{L}_{\mathrm{SB}} = n\ln\left(\left|J\right|\right) + \sum_{i=1}^{n}\ln\left(\left|U_i\right|\right) - 1/2\sum_{i=1}^{n} V_i^2 \tag{4.19}$$

where

$J = \delta\sqrt{G(\delta,\gamma)} / \sqrt{2\pi}$,
$U_i = 1/(H_i + \{1/(\sigma - H_i)\})$,
$H_i = \{(Y_i - \mu)\sqrt{G(\delta,\gamma)}\} + \sigma F(\delta,\gamma)$,
$V_i = \delta \ln(H_i) - \delta \ln(\sigma H_i) + \gamma$,
$G(\delta,\gamma)$ and $F(\delta,\gamma)$ are defined in equation 4.11

and i = 1,…, n refers to the observations.

Beta:

$$L_{Beta} = \sum_{i=1}^{n} \ln\left(\left|\sqrt{G_i}/\sigma\right|\right) + n\ln(\Gamma(\alpha_1 + \alpha_2)) - n\ln(\Gamma(\alpha_1)) - n\ln(\Gamma(\alpha_2)) + (\alpha_1 - 1)\sum_{i=1}^{n}\ln(P_i) + (\alpha_2 - 1)\sum_{i=1}^{n}\ln(1 - P_i) \quad (4.20)$$

where

$$G_i = \left\{\frac{\Gamma(\alpha_1 + 2)\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(2 + \alpha_1 + \alpha_2)}\right\} - \left\{\frac{\alpha_1}{\alpha_1 + \alpha_2}\right\}^2,$$

$$P_i = \left\{\frac{\alpha_1}{\alpha_1 + \alpha_2}\right\} + \left\{(Y_i - \mu)\sqrt{G_i}\right\}/\sigma,$$

and i = 1,…, n refers to the observations.

Using the GAUSS matrix algebra program, the computed maximum likelihood values were then compared among the crop yield distributions examined to assess the fit of the distribution. Since each distribution examined were re-parameterized to have the same mean and variance functions (i.e. they differ only in shape parameters and the higher moments), comparison of likelihood values is a viable procedure to test which distribution fits the data better (even though each distribution is not nested with each other or with another more general distribution). The likelihood value represents the amount of the relationship between the variables that is explained by a model. The higher the likelihood value, the better the model fits the data.

Note that the estimated yield distributions account for any potential non-normality (positive and negative skewness and kurtosis), and heteroskedasticity. After estimating the parameters of the yield distributions, they were used to simulate 10,000 random draws from each of the estimated distribution of corn yields. The lower bound on simulated corn yield was approximated as one standard deviation below the minimum observed yield, which is within the range used in previous studies (see Goodwin and Ker for example).

Actuarially Fair Premium (AFP) Estimation

A description of how the Actuarially Fair Premium (AFP) for Actual Production History (APH) insurance is calculated can be found from several previous studies such as: Skees and Reed (1986), Botts and Boles (1958), Goodwin and Ker (1998), Miller, Kahl and Rathwell (2000a and 200b), Goodwin (1994) and Sherrick, et al. (2004). In order to compute for AFP, some terms need to be defined. First, APH insurance is a yield insurance product covering yield losses from a farm. APH pays an indemnity when the actual yield is below a yield guarantee. Under the Federal Crop Insurance Program, crop producers may select guaranteed yield levels from 50 to 85 percent, in five percent increments of the expected yield. In this context, the yield guarantee is the expected yield multiplied by the farmer-chosen yield election. In mathematical form, let λ_j be the jth coverage level (0.50, 0.55, 0.65 …, 0.85) and let y^e_k be the expected yield of farm *k*. The expected yield of farm *k* can be calculated by taking the average of the simulated yields drawn from the estimated farm-level distribution for that particular farm. Then, the yield guarantee for farm *k* under λ_j coverage level is

$$y^g_{jk} = \lambda_j y^e_k \tag{4.21}$$

and the yield loss valued at crop insurance price election, p_{jk}, for a given yield occurrence (y_{ki}) is

$$L_{jki} = p_{jk} \max\,[y^g_{jk} - y_{ki}, 0]. \tag{4.22}$$

The crop producer may select average price election levels that are determined from FCIC estimates of expected prices. The expected loss for farm k with coverage level λ_j is then computed as

$$L^e_{jk} = \mathrm{E}(L_{jki}) = \sum_{i=1}^{m} (L_{jki}) \,/\, m \tag{4.23}$$

where L_{jki} refers to the losses in equation (4.23) for each of the m simulated yield values. The AFP is then the ratio of the expected loss to the total insured liability, i.e.,

$$\mathrm{AFP}_{jk} = L^e_{jk} \,/\, y^g_{jk} \tag{4.24}$$

The estimated yield distributions from the previous section was used to randomly draw yields (y_{ki} of equation (4.22)) and calculate the AFP on each corn farm considered. Adopting the Sherrick, et al. (2004) approach, AFP was presented in dollars per acre (i.e. L^e_{jk}) and the economic significance of the choice of distribution was measured by calculating the absolute difference of the estimated AFP across the candidate distributions at the 85% coverage level. A price election of 100 percent of the Federal Crop Insurance Corporation (FCIC) expected price was used in the computation.

The computed AFPs using the alternative distributions were also compared to the actual premiums based on RMA's premium rate calculator found in the internet (see http://www3.rma.usda.gov/apps/premcalc/). Note that 10-year historical yields were used for calculating the actual premium, which is consistent with current RMA procedures for premium calculation (see Josephson, Lord and Mitchell (2000) for more details about the current RMA premium rate calculations).

Yield Data

The corn farm-level yield data for this study came from the University of Illinois Endowment Farms. The managers of the Endowment Farms control over 11,000 acres distributed among farms ranging from 40 acres to 1,200 acres. The Endowment Farms are located in twelve counties in Illinois, in approximately 200 miles north to south by 150 miles east to west and are represented in this study by a single letter from "*a*" to "*z*" for identification purposes. The farms are rented to more than forty farm operators predominantly under the common practice of 50-50 share rental arrangements. The manner in which it is operated is similar to commercial operations in Illinois and provides high quality yield data under accurate and consistent recordkeeping practices.

This data were used by Sherrick, et al. (2004) where the farm data was required to have at least twenty observations over the sample period. Resulting from this criterion, the data used in their study were not a random-sample of Illinois farms. In particular, the data consist of twenty six farms with a corn yield series and twenty-five farms with a soybeans series. While their data covered only from 1972-1999, this study used the extended data from 1959 to 2003 but was limited to the corn series only. This data permits estimation of yield models with a high degree of efficiency because it is characterized by relatively long sample period and large number of sites. While parametric approach is recognized to work well in small sample conditions, the use of high quality data set in this study is still pertinent when comparing yield models. Note that since the data series are not a random sample, the results of this study may not be generalizable for the Illinois farm sector as a whole, but rather the results may only represent those farms similar to the non-random Endowment farms used in the analysis. These are farms that tend to have yields that are slightly above average.

Using time series data for yield modeling requires some care due to temporal correlation that affects estimation and inference (Wooldridge, 2003). For this reason, the tests for autocorrelation, heteroskedasticity and unit root were addressed prior to estimating the yield distribution (refer to Table 4.1). With a greater power to reject a null hypothesis of unit root, the Phillips-Peron test was performed to test for the existence of stochastic trend. All twenty-six corn series were found to have a unit root (trend stationary). To test for heteroskedasticity, the special version of the White test was used where the squared residuals are polynomial functions of the fitted values of the dependent variable. The test revealed that farms *k*, *o*, and *t* are found to have statistically significant heteroskedasticity. Finally, the Breush-Godfrey test was used to test for higher order serial correlation. The test indicated low incidence of first-order autocorrelation (only four farms: *e*, *f*, *h*, and *u* had statistically significant autocorrelation).

Based on the tests above, the corn series was de-trended to a base year of 2004 and heteroskedasticity problem was incorporated in the model. Due to low incidence of autocorrelation problem of the sample data, this problem was not explicitly included in the models above.

Table 4.1. Unit Root, Heteroskedasticity and Autocorrelation Tests

Farm	County	Phillips-Peron Test for Unit Root[1]	White test for Heteroskedasticity	Breush-GodfreyTest for Autocorrelation
a	De Kalb	0.0000	0.2783	0.2584
b	De Kalb	0.0000	0.1306	0.2113
c	De Kalb	0.0000	0.1474	0.1143
d	La Salle	0.0000	0.1251	0.2988
e	Wabash	0.0000	0.2218	0.0158**
f	De Witt	0.0000	0.2879	0.0211**
g	De Witt	0.0000	0.2536	0.3180
h	Macon	0.0000	0.9149	0.0830*
i	La Salle	0.0000	0.3322	0.9326
j	Champaign	0.0000	0.3173	0.5068
k	Champaign	0.0060	0.0357**	0.9365
l	Champaign	0.0000	0.6163	0.4766
m	Champaign	0.0000	0.6730	0.1334
n	Douglas, Moultrie	0.0000	0.1304	0.2971
o	Piatt	0.0000	0.0957*	0.9649
p	Piatt	0.0000	0.2664	0.7523
q	Piatt	0.0000	0.1231	0.2027
r	Piatt	0.0000	0.5452	0.4266
s	Piatt	0.0000	0.3047	0.7639
t	Moultrie	0.0000	0.0000***	0.1021
u	Vermilion	0.0000	0.1459	0.0481**
v	Sangamon	0.0000	0.9463	0.8798
w	Sangamon	0.0000	0.5985	0.3193
x	Menard	0.0105	0.2817	0.9378
y	Sangamon, Mccoupin	0.0004	0.4433	0.9645
z	Vermilion	0.0000	0.2217	0.3547

[1]All farms are statistically significant at 1% level.
Note: *** indicates significance at the 1% level; ** indicates significance at the 5% level; and * indicates significance at the 10% level.

Table 4.2 provides the descriptive statistics of the de-trended corn series. The mean of corn yield across farms is 160.39 bu/acre that range from 123.03 bu/acre to184.53 bu/acre. The yield across farms varies with standard deviation of 24.90 ranging from 17.92 bu/acre to 37.11 bu/acre. Except for two farms (farms *l* and *n*), all farms exhibit negative sample skewness. On the other hand, four farms (*f*, *h*, *j* and *n*) exhibit negative kurtosis.

Table 4.2. Yield Data Summaries, University of Illinois Endowment Farms, 1959-2003

Farm	County	Number of Obs.	Mean	Std. Dev.	Skewness	Kurtosis
a	De Kalb	44	170.06	21.17	-1.89	6.64
b	De Kalb	32	139.49	17.92	-2.08	7.53
c	De Kalb	44	150.56	18.21	-0.35	0.29
d	La Salle	43	166.08	22.97	-1.23	3.65
e	Wabash	25	183.19	25.44	-1.22	2.76
f	De Witt	27	163.84	29.37	-0.33	-0.40
g	De Witt	31	178.26	23.05	-1.41	2.37
h	Macon	34	184.53	30.17	-0.24	-0.52
i	La Salle	43	161.68	18.54	-0.72	2.35
j	Champaign	32	171.88	26.17	-0.62	-0.10
k	Champaign	27	145.25	26.62	-0.76	0.80
l	Champaign	29	136.76	23.69	0.01	0.08
m	Champaign	37	168.78	25.84	-0.86	0.76
n	Douglas, Moultrie	45	153.86	20.53	0.07	-0.34
o	Piatt	42	160.07	25.80	-0.66	0.06
p	Piatt	42	154.40	28.04	-0.79	0.44
q	Piatt	40	176.35	21.70	-0.99	0.68
r	Piatt	33	146.91	23.46	-1.16	1.40
s	Piatt	40	160.32	24.79	-0.64	0.99
t	Moultrie	29	157.78	25.04	-0.68	0.45
u	Vermilion	44	128.25	25.75	-0.72	0.68
v	Sangamon	29	174.39	23.64	-1.17	2.95
w	Sangamon	29	181.37	23.87	-0.96	3.00
x	Menard	20	157.38	37.11	-1.64	2.09
y	Sangamon, Mccoupin	29	175.73	27.66	-0.82	1.30
z	Vermilion	30	123.03	30.83	-0.50	0.04
	Average	35	160.39	24.90	-0.86	1.54
	Minimum	20	123.03	17.92	-2.08	-0.52
	Maximum	45	184.53	37.11	0.07	7.53

Note: The yield data was de-trended to a base year 2004.

In summary, the variety of yield crop behavior is presented in figure 4.2. The figure shows that the corn yield distribution can be significantly skewed either to the right or left and it could either be platykurtic or leptokurtic. Note that these empirical distributions are based on the de-trended corn yield data.

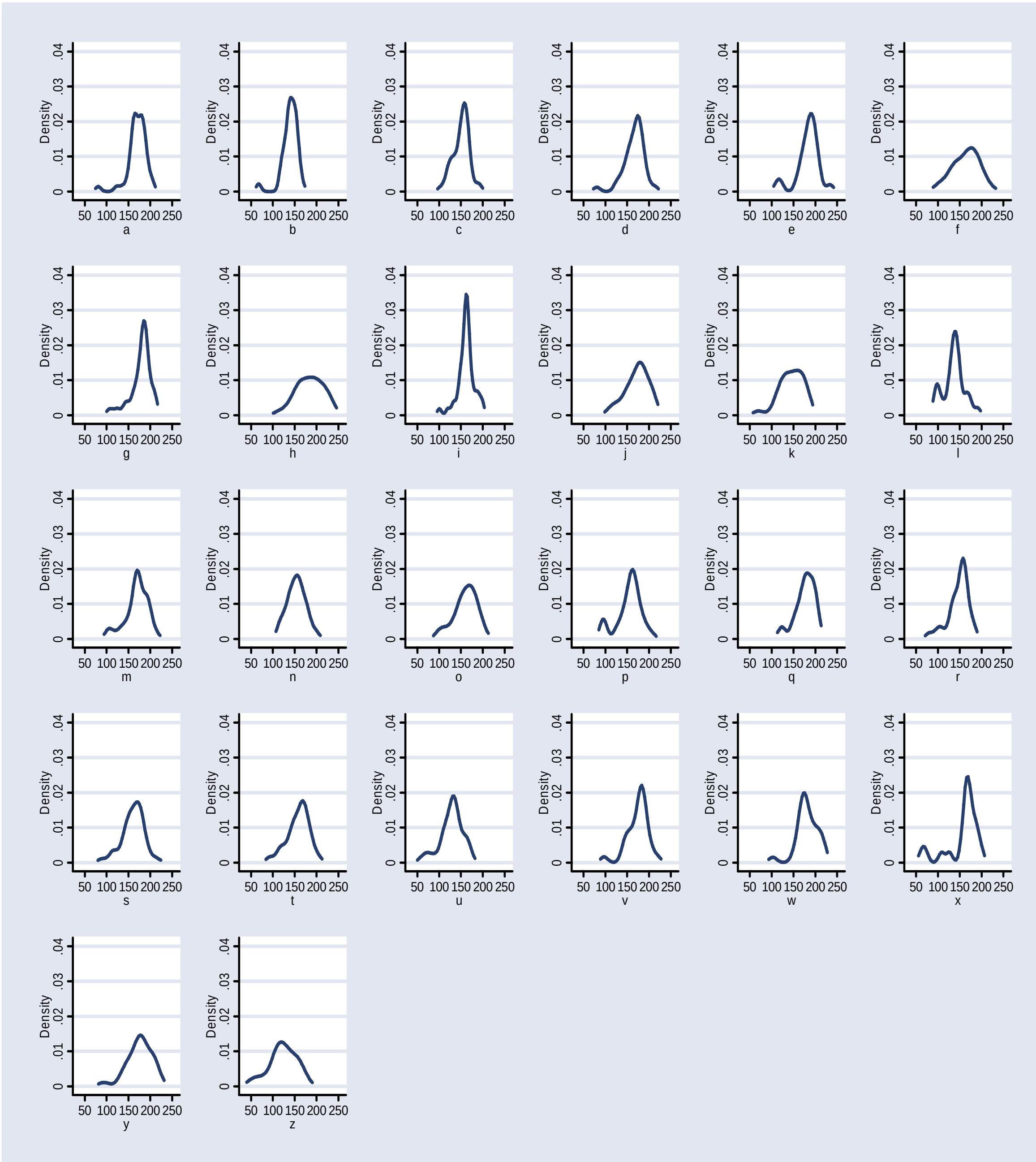


Figure 4.2. Empirical Probability Distribution Functions, De-trended Illinois Corn Yield, 1959-2003

CHAPTER V

RESULTS

The first part of this chapter presents the estimated parameters of all the alternative parametric yield distributions considered in this study. The statistical performance of the Johnson family of distributions in modeling crop yields is then compared relative to the beta distribution in the second part of this section. Finally, the last section examines the APH premium rate effects of using the different distributions considered in the study. This determines the economic impact of using alternative yield distributions in crop insurance premium rate setting.

Estimated Corn Yield Models

Maximum likelihood estimation was used to estimate the parameters of the alternative distributions to model yields. Starting values used for each iterations were selected so that the values gave a proper convergence to achieve a global maximum (i.e. the gradient is zero). The parameter estimates and related statistics for four models are presented in Tables 5.1 to 5.4.

If one specific farm is examined, say farm *f*, it can be seen that the predicted behavior of corn yield depends on the choice of distribution used. Under the lognormal model, the mean yields are not a statistically significant function of time (based on the estimated coefficients of β_1 and β_2). However, under the three other distributions, (S_U, S_B, and beta), βs are significant at the 10% level of significance. On average, the variance of corn yield, as measured by its standard deviation, σ_1, appear to be changing through time under the S_B and beta distributions, but it does not seem to be significantly changing through time under the S_U and lognormal distributions (on average).

Generally, the δ parameter under the lognormal distribution is expected to be large since it tends to approach the normal distribution. Based on the non-significance of shape parameters for S_U (i.e., μ and θ), farm *f* is not normally distributed. The non-significance of the skewness parameter μ implies that the distribution of yield for farm *f* is symmetric. Likewise, a positive coefficient of θ suggests a leptokurtic shape of corn yield for farm *f*. In general, if $\theta > 0$ and μ approaches 0 then the yield distribution is symmetric but remains kurtotic. If $\theta > 0$ and $\mu > 0$, the yield distribution is kurtotic and right skewed distribution. If $\theta > 0$ and $\mu < 0$, the yield distribution is kurtotic and left skewed distribution (which is consistent with Ramirez, 2000). Note that the behavior of kurtosis and skewness are not directly related to the signs of the shape parameters of S_B (γ and δ) and beta (α_1 and α_2). Recall that from the previous chapter, the parameters δ, α_1 and α_2 are always positive.

Under the lognormal distribution, only five farms have corn yields that have a significant β_2 coefficient (mean parameter) and only fourteen farms have a corn yield variance that changes through time (significant σ_1). Conversely, if the beta, S_B and S_U distributions are employed, most of the farms have a mean and variance of corn yields that are functions of time. In particular, the number of significant parameters for the mean, βs, and variance, σs, are relatively the same among the beta, S_B and S_U distributions. While the difference in range and values of the coefficients for the mean and variance parameters are small, there are discrepancies in the skewness and kurtosis parameters across distributions. In general, this confirms the result of Sherrick, et al. (2004) that the differences

Table 5.1. Parameter Estimates under Lognormal Distribution

Farm	β_0	β_1	β_2	σ_0	σ_1	δ
A	91.760***	2.758***	-2.501	13.986***	0.246***	47.631
b	97.614***	1.500	-1.467	14.340***	0.149***	80.000
c	87.575***	1.332	0.289	12.906***	0.216***	15.629
d	102.998***	0.647	2.362	15.968***	0.264***	74.682
e	64.890***	3.018	2.661	12.984***	0.750***	34.444
f	126.229***	2.336	-4.100	27.998	0.061	80.000
g	121.245***	2.657	-3.018	27.056***	-0.260***	70.652
h	69.581***	5.983***	-8.929***	27.416***	0.051	0.000
i	97.789***	0.740	2.081	13.444***	0.230***	50.801
j	124.784***	2.146	-2.568	29.394***	-0.216**	80.000
k	130.875***	0.943	-1.473	24.820***	0.083	69.659
l	137.897***	0.104***	-0.094***	21.631***	0.012	18.605
m	127.139***	0.890	0.749	21.748***	0.201	80.000
n	94.616***	1.792**	-1.198	15.124***	0.197***	80.000
o	115.975***	1.093***	0.000	20.988***	0.198***	31.059
p	100.054***	0.439	2.523	20.523***	0.316***	62.588
q	131.128***	0.377	2.017	17.385***	0.166***	79.124
r	117.175***	1.271	-1.719	21.026***	0.116	79.987
s	124.471***	1.521**	-1.833**	21.632***	0.127**	80.000
t	141.667***	1.072	-1.880	23.333***	0.083	79.033
u	83.616***	1.024***	-0.036***	19.837***	0.239***	80.000
v	132.960***	0.000	5.874***	22.158***	0.060	55.668
w	149.083***	0.238	3.549	24.731**	-0.078	80.000
x	112.143***	0.000	9.540	29.979***	0.545	80.000
y	130.891***	1.542**	0.022	25.268**	0.142	80.000
z	92.147***	0.877	0.856	34.908***	-0.313	90.000
No. of signif. parameter	26	8	5	25	14	0
Average	111.781	1.396	0.066	21.561	0.138	63.829
Minimum	64.890	0.000	-8.929	12.906	-0.313	0.000
Maximum	149.083	5.983	9.540	34.908	0.750	90.000

Note: *** indicates significant at the 1% level; ** indicates significant at the 5% level; and * indicates significant at the 10% level.

Table 5.2. Parameter Estimates under S_U Distribution

Farm	β_0	β_1	β_2	σ_0	σ_1	μ	θ
a	86.787***	3.133***	-2.989**	14.972***	0.304**	-0.608	0.981***
b	97.279***	0.614***	2.884***	14.009***	0.228***	-6.587	0.742***
c	85.047***	1.422***	0.405	12.531***	0.237***	-17.153***	0.261
d	97.750***	1.319*	0.885	15.628***	0.274***	-1.558	0.492
e	63.316***	2.552***	6.531***	12.241***	0.843***	-10.144	0.518***
f	122.866***	2.794*	-5.271*	27.589**	0.119	-10.000	0.241
g	113.761***	3.790***	-6.213***	25.366***	-0.140***	-8.669	0.571***
h	69.112***	6.028***	-8.997***	25.510***	0.174	-10.000	0.204
i	96.958***	0.961*	1.470	19.425*	0.324*	-0.026	1.285***
j	114.638***	3.186***	-4.755***	35.006***	-0.236***	-8.361	0.611*
k	116.548***	2.291***	-3.579***	33.571***	0.306**	-5.447	0.887***
l	140.489***	-0.010***	0.035***	23.399***	0.000***	3.508	0.000
m	130.379***	0.317*	2.357***	21.827***	0.204***	-13.360	0.350**
n	93.952***	1.857**	-1.313	15.057***	0.200***	-10.000	0.053
o	100.573***	2.764***	-3.330***	21.858***	0.286***	-9.872***	0.553***
p	93.279***	1.307	0.589	20.202***	0.339***	-6.370	0.317
q	120.955***	1.771***	-1.263	18.212***	0.184***	-10.927***	0.469**
r	111.307***	1.815***	-2.630***	21.131***	0.163**	-1.685	0.642
s	120.123***	1.834***	-2.210***	21.729***	0.159***	-0.984	0.613*
t	131.454***	1.605***	-1.163***	26.385***	0.251***	-8.056***	0.631***
u	77.628***	1.786**	-1.674	19.786***	0.256***	-14.707	0.302**
v	135.412***	-0.720	8.613***	21.495**	0.308**	-8.621***	0.598**
w	143.286***	1.878	-2.547	25.241	-0.048	0.168	0.782
x	89.609**	4.476	-10.074	53.924	-0.901	-5.010	0.750
y	123.707***	2.982	-4.877	25.323**	0.157	-10.000	0.324
z	86.162***	2.213***	-3.750***	35.441***	-0.311***	-18.400***	0.246**
No. of signif. parameter	26	21	16	24	21	6	15
Average	106.245	2.075	-1.649	23.341	0.142	-7.418	0.516
Minimum	63.316	-0.720	-10.074	12.241	-0.901	-18.400	0.000
Maximum	143.286	6.028	8.613	53.924	0.843	3.508	1.285

Note: *** indicates significant at the 1% level; ** indicates significant at the 5% level; and
* indicates significant at the 10% level.

Table 5.3. Parameter Estimates under S_B Distribution

Farm	β_0	β_1	β_2	σ_0	σ_1	γ	δ
a	89.791***	3.018***	-3.080*	13.016***	0.233***	-8.700	3.021***
b	97.300***	0.613	2.890	13.885**	0.226**	-7.136***	1.333***
c	86.195***	1.441***	0.126	12.391***	0.215***	-5.556***	7.649***
d	96.735***	1.286*	1.186	15.182***	0.285***	-8.700	2.988
e	63.075***	2.500**	6.991	11.832***	0.833***	-1.718	1.088
f	111.229***	4.130***	-7.793***	25.552***	0.439***	-0.329	0.629**
g	106.325***	4.441***	-7.280***	31.130***	-0.373***	83.997	1.152***
h	70.610***	5.541***	-7.297*	34.103***	-0.218	-0.278	0.679***
i	94.654***	1.231	0.872	13.300***	0.228***	-8.700	5.019*
j	114.090***	3.847***	-5.741***	33.896***	-0.345***	-0.757***	0.391***
k	116.980***	2.342***	-3.659***	24.873***	0.231***	-0.936**	0.602***
l	139.756***	0.058***	-0.101***	23.577***	0.004***	0.272	1.597
m	130.821***	0.244	2.560	21.584***	0.196	-2.066	1.579
n	92.335***	1.344***	0.706	15.367***	0.279***	-0.089	0.624***
o	99.606***	2.846***	-3.429***	19.633***	0.267***	-0.958*	0.868***
p	92.699***	1.264***	0.854***	19.665***	0.347***	-1.665***	1.574***
q	141.744***	-0.376***	2.524***	23.722***	0.125***	-0.344***	0.388***
r	111.340***	1.822***	-2.641***	19.998***	0.155**	-8.000***	2.191***
s	119.658***	1.877***	-2.261***	20.672***	0.155***	-8.700	3.696**
t	141.196***	1.119	-1.963	22.932***	0.086	-2.039	11.705***
u	79.422***	1.488*	-0.961	19.864***	0.260***	-0.767	0.976**
v	134.205***	-0.501***	7.984***	22.122***	0.321***	-0.848***	0.549***
w	141.418***	1.721***	-1.393***	20.630***	0.190***	-8.700	3.390**
x	91.261***	4.784**	12.267**	44.812**	-0.706	-1.044	0.663
y	123.861***	2.876	-4.396	26.918***	0.058	-1.635	1.358
z	47.954***	5.636***	-9.553***	50.000	-1.000	-0.312**	0.622***
No. of signif. parameter	26	21	16	25	20	10	20
Average	105.164	2.177	-0.869	23.102	0.096	0.165	2.167
Minimum	47.954	-0.501	-9.553	11.832	-1.000	-8.700	0.388
Maximum	141.744	5.636	12.267	50.000	0.833	83.997	11.705

Note: *** indicates significant at the 1% level; ** indicates significant at the 5% level; and * indicates significant at the 10% level.

Table 5.4. Parameter Estimates under Beta Distribution

Farm	β_0	β_1	β_2	σ_0	σ_1	α_1	α_2
a	91.997***	2.541***	-1.885	12.354***	0.225***	32.488***	62.203***
b	96.844***	1.543	-1.532	14.044***	0.150***	67.768*	53.829***
c	86.834***	1.371***	0.269***	12.406***	0.213***	57.916	36.056***
d	96.565***	1.240	1.381	15.057***	0.290***	94.564	4.980
e	63.168***	3.069	3.268	11.031***	0.693***	55.359	7.921***
f	121.390***	3.236***	-5.478*	24.813***	0.344***	1.601***	0.991***
g	108.623***	4.330***	-7.098***	30.000***	-0.350***	63.891	1.117***
h	69.005***	6.039***	-9.013***	26.253***	0.129	78.242***	7.669
i	94.739***	1.210	0.930	13.315***	0.229***	89.293	15.315
j	114.731***	3.433***	-5.124***	30.000	-0.050	69.366***	1.016***
k	124.363***	1.613*	-2.521*	21.695***	0.131*	73.586	5.673***
l	140.554***	-0.049	0.162	23.221***	0.002	49.918	43.117
m	131.262***	0.182	2.715	21.819***	0.191	10.440	2.358
n	93.660***	1.831*	-1.201	14.742***	0.200***	5.717***	5.564***
o	100.080***	2.830***	-3.409***	20.566***	0.276***	47.588***	1.773***
p	92.790***	1.256	0.867	19.770***	0.348***	20.399***	3.565***
q	124.277***	1.407	-0.485	16.454***	0.158***	72.416	3.727***
r	111.808***	1.778***	-2.576***	19.652***	0.148**	95.513***	3.005***
s	119.772***	1.867***	-2.249***	20.634***	0.154***	92.542***	8.825**
t	139.226***	1.307	-2.264	22.080***	0.099	75.305	21.953***
u	77.780***	1.735*	-1.511	19.806***	0.263***	87.421	3.795
v	135.239***	-0.570***	8.434***	19.504***	0.291***	88.958***	1.309***
w	140.678***	1.817	-1.617	20.859***	0.197**	63.905***	4.582
x	113.054***	0.173***	8.331***	26.821***	0.463	93.075***	2.207***
y	125.549***	2.779	-4.405	22.690**	0.245	94.304**	5.495
z	83.520***	2.695***	-4.296***	29.887***	0.138	1.643***	0.939***
No. of signif. parameter	26	15	12	26	18	15	18
Average	107.597	1.948	-1.166	20.364	0.199	60.893	11.884
Minimum	63.168	-0.570	-9.013	11.031	-0.350	1.601	0.939
Maximum	140.678	6.039	8.434	30.000	0.693	95.513	62.203

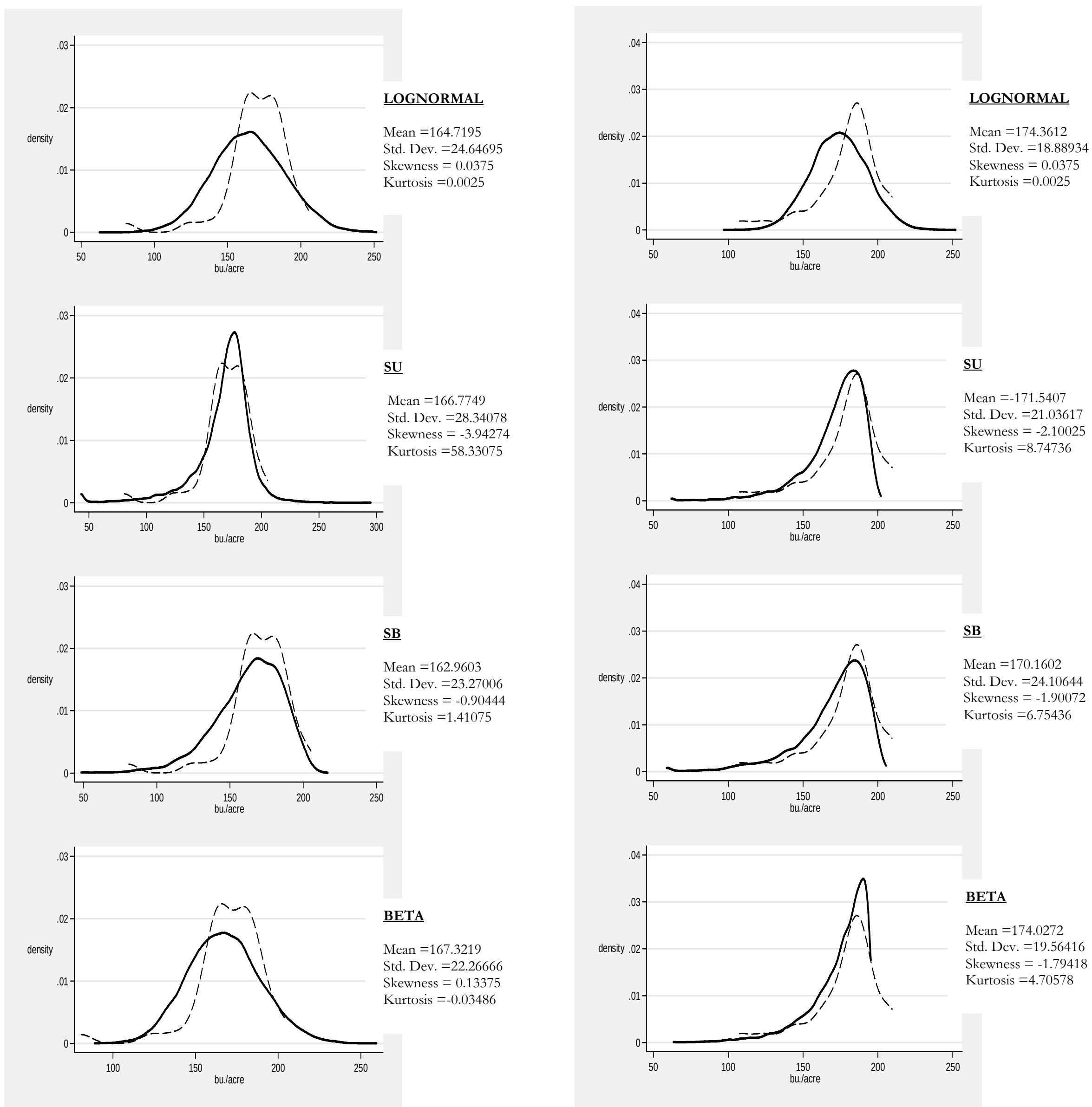
Note: *** indicates significant at the 1% level; ** indicates significant at the 5% level; and * indicates significant at the 10% level.

in the parameter estimates depend on the underlying characteristics of each distribution. However, the real question is which distribution more accurately depicts actual crop yield behavior. To address this concern, we assessed the statistical performance of the distributions by comparing the likelihood values from each distribution in the next section.

Comparison of the Statistical Performance of the Alternative Distributions: Johnson Family vs. Beta

As an initial step in comparing the Johnson family of distributions against the beta, we visually compared the estimated probability density function (pdf) of the different parametric distributions relative to the actual empirical distribution of the de-trended yields. For the purpose of illustration, only farms *a*, and *g* are considered in the discussion of results below. In Figure 5.1, the dashed line represents the empirical distribution of the de-trended corn yields and the solid line depicts the pdf from the candidate distributions. In Panel A of figure 5.1, the empirical distribution appears to be bimodal. With some crops potentially having a bimodal distribution, the normal and the beta distribution have been criticized because of their unimodal yield distribution (Goodwin and Ker, 1998 and Miller, Kahl and Ratwell, 2000). From visual evidence, lognormal and beta distribution by far could not represent the sample data in farm *a* as compared with the S_U and S_B distributions. S_U seems to fit the data and appears to be a flexible model even if the empirical distribution is bimodal and the S_U is unimodal.

A. Probability distribution functions, farm *a,* B. Probability distribution functions, farm *g,*



Legend: ------ Empirical Distribution —— Candidate Distribution (Lognormal, S_U, S_B, Beta)

Figure 5.1. Probability Distribution Functions and Empirical Distribution for Illinois Corn Yields, 1959-2003

While it is fairly obvious in panel A of figure 5.1 which distribution seem to best represent the sample data, it is not the case for the figure in panel B. Since the empirical data of farm *g* is characterized by negatively skewed yield behavior, it is expected that the lognormal distribution will fit the data poorly (because the lognormal does not allow for negative skewness). The S_U, lognormal, and the beta distributions appear to represent the sample data fairly well, however, it is not visually clear which among the three distributions fits the data best. In this scenario, visual evidence may not be sufficient to support any statement regarding the fit of the S_U, lognormal, and the beta distribution to the yield behavior of farm *g*. Thus, the use of statistical tools must supplement the visual evidence. Thus, comparison of likelihood values is used to provide statistical evidence on the statistical performance of the alternative distributions and gauge which parametric distribution fits the data best. The likelihood value is computed for each farm and these values are presented in Table 5.5.

Based on the likelihood values, the lognormal distribution remains the poorest performer for all farms. Again, this is primarily because the lognormal distribution does not permit negative skewness and kurtosis. On the other hand, the S_U, S_B and beta gave relatively close likelihood values since these distributions permit negative skewness. The S_U distribution fits the data best when kurtosis has a large positive value. This is consistent with the figure from the previous section as described by Ramirez and McDonald (2005). But note that the skewness-kurtosis plane for the S_U distribution is bounded by positive kurtosis.

On the other hand, the likelihood values for the S_B and the beta distributions are very similar. The computed skewness across all farms is not significantly different at any level of significance using the Wilcoxon-matched-paired signed rank test. This is also consistent with the figure by Ramirez and McDonald (forthcoming) wherein the skewness-kurtosis plane of the beta distribution is a subset of the S_B distribution. Similarly, they outperform the S_U when the distribution of the empirical data has a negative kurtosis. Hence, for yields that actually have extreme negative or positive skewness, together with negative kurtosis, the S_B and beta distributions give very similar likelihood values. This suggests that it is important that assumed crop yield distributions permit not only negative skewness but also negative kurtosis.

From the computed likelihood values, the S_B distribution is the best fitting yield distribution in the majority of the Illinois farms considered in the study. Among the 26 farms, 17 farms had the S_B distribution as the best distribution that fits the actual yield behavior in those farms. It is interesting to note that S_B consistently fits the yield distribution when the observed yield data has negative kurtosis (see farms *f*, *h*, *j* and *n*; in table 4.2). Likewise, note for the two cases where the beta distribution outperforms the S_B distribution (farms *m* and *w*), there is only a slight difference in the magnitude of the likelihood value (less than 0.1). Overall, the likelihood analysis shows that if the lognormal, S_U and S_B are considered as one model, this has the potential to be a highly flexible yield model that permits a wider range of possible combinations of skewness and kurtosis.

Table 5.5. Likelihood Value Comparisons, Skewness and Kurtosis of the Johnson Family and the Beta Distributions

Farm	Lognormal			S_U			S_B			Beta		
	Likelihood	Skewness	Kurtosis	Likelihood	Skewness	Kurtosis	Likelihood	Skewness	Kurtosis	Likelihood	Skewness	Kurtosis
a	-384.78	0.0630	0.0071	-367.24	-3.94	58.33	-373.34	-0.90	1.41	-392.59	0.13	-0.03
b	-270.70	0.0375	0.0025	-247.61	-3.20	22.51	-247.61	-3.05	19.19	-269.72	-0.04	-0.05
c	-376.55	0.1924	0.0659	-372.75	-0.81	1.20	-374.37	-0.13	0.00	-374.49	-0.10	-0.05
d	-385.71	0.0402	0.0029	-378.45	-1.17	3.36	-378.77	-0.92	1.46	-379.08	-0.82	0.93
e	-225.46	0.0871	0.0135	-216.17	-1.84	6.54	-216.00	-1.17	1.28	-219.44	-0.56	0.37
f	-258.11	0.0375	0.0025	-256.63	-0.74	0.99	-254.16	-0.34	-1.07	-254.89	-0.40	-0.59
g	-281.20	0.0425	0.0032	-267.11	-2.10	8.75	-265.64	-1.79	4.71	-265.67	-1.80	4.71
h	-323.73	0.0300	0.0016	-322.30	-0.61	0.68	-319.25	-0.28	-1.07	-322.05	-0.61	0.49
i	-371.76	0.0591	0.0062	-362.53	-0.83	396.69	-369.68	-0.41	0.26	-369.87	-0.39	0.17
j	-298.12	0.0375	0.0025	-291.79	-2.32	10.94	-276.99	-1.05	-0.23	-285.35	-1.90	5.26
k	-252.57	0.0431	0.0033	-241.47	-4.58	52.76	-237.32	-1.06	0.18	-246.88	-0.73	0.72
l	-265.82	0.1615	0.0464	-265.11	0.00	0.00	-265.01	0.11	-0.49	-265.10	-0.03	-0.06
m	-344.35	0.0375	0.0025	-338.16	-1.13	2.36	-337.99	-0.84	0.67	-337.91	-0.82	0.56
n	-394.97	0.0375	0.0025	-394.91	-0.08	0.02	-390.30	-0.09	-1.21	-393.73	-0.01	-0.42
o	-389.78	0.0967	0.0166	-379.07	-2.01	7.93	-376.80	-0.80	-0.07	-377.76	-1.38	2.68
p	-396.02	0.0479	0.0041	-390.69	-0.98	1.78	-390.57	-0.68	0.26	-390.62	-0.76	0.61
q	-356.76	0.0379	0.0026	-348.13	-1.61	4.92	-337.56	-0.46	-1.27	-348.35	-0.94	1.23
r	-300.69	0.0375	0.0025	-290.72	-2.10	10.01	-290.94	-1.40	3.54	-291.34	-1.08	1.69
s	-368.64	0.0375	0.0025	-363.53	-1.40	5.73	-364.69	-0.67	0.74	-364.99	-0.57	0.43
t	-267.85	0.0380	0.0026	-262.14	-2.44	12.17	-267.43	-0.03	7.39	-265.88	-0.26	0.04
u	-408.38	0.0375	0.0025	-403.65	-0.96	1.67	-402.42	-0.57	-0.42	-403.14	-0.95	1.26
v	-263.98	0.0539	0.0052	-255.56	-2.25	10.17	-252.39	-1.01	-0.05	-252.72	-1.68	4.13
w	-265.39	0.0375	0.0025	-262.53	0.58	8.22	-262.47	-0.76	0.98	-262.40	-0.82	0.91
x	-199.42	0.0375	0.0025	-188.19	-3.26	23.56	-187.08	-1.12	0.45	-190.28	-1.28	2.37
y	-274.08	0.0375	0.0025	-270.27	-1.03	1.95	-269.99	-0.83	0.47	-270.22	-0.77	0.82
z	-289.01	0.0333	0.0020	-287.43	-0.76	1.06	-285.15	-0.33	-1.09	-287.49	-0.75	0.75
No. of times that model ranked first		0			7			17			2	

Note: Shaded areas are the highest likelihood value for each farm.

Impacts on the Actual Production History (APH) Insurance Premium Rates

From the previous section, we showed that more flexible parametric distributions may provide a better fit of the actual crop yield behavior for some Illinois corn farms. In this section, we discuss the potential economic impact of using the more flexible Johnson family of distributions by examining the effect of using these distributions to set the premium rates in the Actual Production History (APH) insurance program. While corn producers in Illinois can choose among APH yield insurance, revenue insurance products (Crop Revenue Coverage, Revenue Assurance, Income Protection) and area yield insurance (Group Risk Protection and Group Risk Income Protection), this study limits the analysis to the APH program due to its popularity in the area and its simple contract form. This type of insurance is fairly straightforward yield insurance contract that allows one to tractably evaluate the impact of modeling yields based on different parametric distributional assumptions (Sherrick, et al., 2004). For example, the APH contract form would not necessitate the calculation of joint yield-price distributions which would be required in an evaluation of revenue insurance contracts. Before proceeding to our results in this section, one important caveat to keep in mind is that the farms used in the analysis are not a random sample that could represent all farms in the region. The yield histories of the farms investigated in this study tend to be slightly above the average. Hence, the results may only be generalized to farms that tend to be above average.

Using the APH formula from the previous chapter, the premiums calculated under different distributional assumptions, as well as the actual premiums based on the RMA premium rate calculator, are presented in Table 5.6. Note that the RMA premium rates presented here are unsubsidized rates. Also, for easier comparison, the premium rates presented below are in $ per acre rather than $ per unit of liability (as shown in equation 4.24). The corresponding AFPs in $ per unit of liability are shown in the Appendix. As shown in the table, the actuarially fair premium (AFP) of farm *a* is $5.02 per acre under the

Table 5.6. Comparison of Calculated AFPs from the Alternative Distributions vs. the Actual Unsubsidized APH Premium Rate at the 85% Coverage Level and 100% Price Election ($ per Acre)

Farm	County	Lognormal	S_U	S_B	Beta	Actual Premium
a	De Kalb	5.02	7.34	5.78	3.22	12.00
b	De Kalb	3.76	5.73	6.30	3.61	14.00
c	De Kalb	4.01	6.05	4.41	4.44	13.00
d	La Salle	5.83	8.47	7.75	7.83	13.00
e	Wabash	10.11	12.83	13.24	8.74	18.00
f	De Witt	8.37	11.43	16.68	12.91	11.00
g	De Witt	1.74	5.23	4.56	4.35	11.00
h	Macon	7.64	10.79	6.58	10.08	12.00
i	La Salle	3.98	5.94	5.28	4.98	13.00
j	Champaign	3.34	8.41	5.78	10.11	12.00
k	Champaign	8.03	16.77	16.01	8.32	11.00
l	Champaign	4.40	5.64	5.57	5.66	11.00
m	Champaign	7.29	9.56	9.18	9.10	12.00
n	Douglas, Moultrie	5.51	5.68	7.49	5.19	12.00
o	Piatt	7.86	13.85	12.70	13.25	11.00
p	Piatt	11.18	13.80	14.33	14.15	11.00
q	Piatt	3.84	7.66	9.70	5.32	12.00
r	Piatt	7.00	9.40	9.13	8.80	11.00
s	Piatt	7.20	9.02	8.47	8.28	11.00
t	Moultrie	6.18	12.59	6.16	6.22	11.00
u	Vermilion	11.87	13.84	15.22	15.18	14.00
v	Sangamon	3.41	9.35	12.49	8.07	12.00
w	Sangamon	2.63	3.25	6.84	6.78	12.00
x	Menard	17.90	15.42	15.62	16.83	13.00
y	Sangamon, Mccoupin	7.30	10.09	9.35	9.88	12.00
z	Vermilion	8.33	10.67	5.09	13.08	14.00
	Average	6.68	9.57	9.22	8.63	12.27
	Std. Dev.	3.46	3.51	3.97	3.75	1.54
	Minimum	1.74	3.25	4.41	3.22	11.00
	Maximum	17.90	16.77	16.68	16.83	18.00

Notes: AFP computation is adopted from Sherrick's et al. paper where AFP is equal to expected payouts in dollar per acre. Refer to Appendix for AFP $ per unit of liability.

lognormal distribution. The highest AFP generated is $7.34 under the S_U distribution (which is the best fitting distribution based on our likelihood value comparisons above). In general, the beta distribution generates higher actuarially fair premiums than the lognormal distribution, which is consistent with Sherrick et al.'s (2004) findings. However, in comparison with S_B and S_U distributions, the beta distribution generated lower AFPs on average. In seven cases under S_B and S_U distributions and five cases under the beta distribution, the computed AFP is larger than the existing premium. On average, the existing unsubsidized premium based on the current RMA rate-making procedures tend to be higher than the estimated premiums based on the alternative parametric distributions. In summary, the results above suggest that: (1) premium rates calculated based on the beta distribution tend to be lower relative to Johnson distributions (S_B and S_U), and (2) the actual unsubsidized premium rates tend to be higher than the AFPs based on the Johnson distributions. In general, these results indicate that the implied valuation of crop insurance is sensitive to the choice of distributional specification.

To further analyze the differences in computed AFP, the mean absolute-percentage difference among the alternative distributions (at the 85% coverage level) is summarized in table 5.7. For example, the mean absolute-percentage differences between AFP calculated based on S_U, S_B and beta against lognormal are 32.31%, 32.27% and 27.01% respectively. This implies that AFP calculated using S_U distribution is 32.31% greater than the AFP calculated using lognormal distribution. Similarly, the premium based on the S_B distribution is 32.27% greater than premium based on lognormal distribution. Hence, the AFP generated using lognormal distribution is underestimated (32% to 49% difference) compared to other candidate distributions and the actual premium rates. Similarly, the AFPs generated from S_U distribution and S_B distribution tend to be higher than beta distribution (by 28.1% and 26.46% in magnitude, respectively). This shows that AFP is underestimated when the beta distribution is used to calculate the AFP rather than S_B and S_U.

Table 5.7. Mean Absolute-Percentage Difference in Actuarially Fair Premiums

	Lognormal	S_U	S_B	Beta	Actual Premium
Lognormal	-	32.31%***	32.27%***	27.01%***	48.49%***
S_U		-	(24.67)%	(28.61)%*	31.87%***
S_B			-	(26.46)%	37.42%***
Beta				-	36.95%***

Note: *** indicates significant at a 1% ; *indicates significant at a 10% level under Wilcoxon signed ranks test; and a parenthesis indicates a negative difference.

Using pairwise Wilcoxon-signed rank test, the statistical significance of the absolute-percentage differences was also tested. The actual unsubsidized premiums based on current RMA rate-making procedures tend to be statistically different from all the AFPs computed based on the alternative distributions. Only the percentage difference between S_U and S_B, as well as beta and S_B, are not statistically significant.

In light of the statistical difference between the actual unsubsidized premiums and the AFPs calculated using the candidate distributions, the correlation coefficient for each of the AFP under each distribution versus the actual unsubsidized premium was computed. In general, the actual premium and the distribution-based AFPs have low correlations with values equal to 0.3466 for lognormal, 0.1446 for S_U, 0.1411 for S_B and 0.1973 for beta (consistent with Sherrick et al's.(2004) findings). Low correlation between the distribution-based AFPs and the actual premiums is suggestive of adverse selection in the crop insurance program (Sherrick, et al., 2004).

The difference between the AFPs based on the Johnson family of distributions and the actual unsubsidized premium may suggest that the current RMA premium estimation may have failed to accurately reflect the individuals' probability of loss. This, in turn, could potentially cause adverse selection in the program. As mentioned in the conceptual framework, the higher unsubsidized actual premium rate may discourage low risk farmers from participating in the insurance program (Knight and Coble, 1997). However, note that the AFPs based on the Johnson family of distributions are being compared with the unsubsidized actual premiums. Hence, to more fairly assess the presence of adverse selection in the crop insurance program, one should also compare the distribution-based AFPs with the subsidized premium rates.

In Table 5.8, we compare the AFPs based on the best-fitting distribution versus the actual premium rate when the current subsidy level at the 85% coverage level is taken into consideration. As authorized by the Agricultural Risk Protection Act of 2000, the current subsidy level for Illinois corn farm at the 85% coverage level is 38% (Glauber, 2004). When the current subsidy level is taken into account, the magnitude of the difference between the AFP from the best-fitting distribution versus the actual premium becomes smaller. On average, the distribution-based AFPs and the actual subsidized premiums are closer in value. This result suggests that the subsidy may have acted as a tool to reduce the adverse selection in the program. For example, without the subsidy, the difference between the distribution based premium and the actual premium for farm *a* is around $4.66. But if the subsidy is considered, the difference is reduced to about $0.10. Hence, it seems that the subsidies implemented to encourage participation of low-risk producers also tend to reduce the adverse selection present in the program.

To further show that premiums generated based on the best-fitting distribution are relatively close to the subsidized premiums, the mean absolute-percentage differences among

Table 5.8. Comparison of AFPs based on the Best-Fitting Distribution vs. the Actual Subsidized and Unsubsidized Premium Rates

Farm	County	Premium based on the Best-Fitting Distribution[1] (A)	Unsubsidized Actual Premium (B)	Subsidized Actual Premium[2] (C)	B-A[3]	C-A[4]
a	De Kalb	7.34[SU]	12.00	7.44	4.66	0.10
b	De Kalb	6.30	14.00	8.68	7.70	2.38
c	De Kalb	6.05[SU]	13.00	8.06	6.95	2.01
d	La Salle	8.47[SU]	13.00	8.06	4.53	-0.41
e	Wabash	13.24	18.00	11.16	4.76	-2.08
f	De Witt	16.68	11.00	6.82	-5.68	-9.86
g	De Witt	4.56	11.00	6.82	6.44	2.26
h	Macon	6.58	12.00	7.44	5.42	0.86
i	La Salle	5.94[SU]	13.00	8.06	7.06	2.12
j	Champaign	5.78	12.00	7.44	6.22	1.66
k	Champaign	16.01	11.00	6.82	-5.01	-9.19
l	Champaign	5.57	11.00	6.82	5.43	1.25
m	Champaign	9.10[beta]	12.00	7.44	2.90	-1.66
n	Douglas, Moultrie	7.49	12.00	7.44	4.51	-0.05
o	Piatt	12.70	11.00	6.82	-1.70	-5.88
p	Piatt	14.33	11.00	6.82	-3.33	-7.51
q	Piatt	9.70	12.00	7.44	2.30	-2.26
r	Piatt	9.40[SU]	11.00	6.82	1.60	-2.58
s	Piatt	9.02[SU]	11.00	6.82	1.98	-2.20
t	Moultrie	12.59[SU]	11.00	6.82	-1.59	-5.77
u	Vermilion	15.22	14.00	8.68	-1.22	-6.54
v	Sangamon	12.49	12.00	7.44	-0.49	-5.05
w	Sangamon	6.78[beta]	12.00	7.44	5.22	0.66
x	Menard	15.62	13.00	8.06	-2.62	-7.56
y	Sangamon, Mccoupin	9.35	12.00	7.44	2.65	-1.91
z	Vermilion	5.09	14.00	8.68	8.91	3.59
	Average	9.67	12.27	7.61	2.60	-2.06
	Std. Dev.	3.81	1.54	0.95	4.11	3.93
	Minimum	4.56	11.00	6.82	-5.68	-9.86
	Maximum	16.68	18.00	11.16	8.91	3.59

[1]SU indicates that S_u distribution is best-fitting model and beta indicates that beta was the best-fitting model; S_B otherwise.
[2]This is actual premium less the 38% subsidy at the 85% coverage level.
[3]This is the difference between the premiums from the best fitting distribution and the unsubsidized actual premium.
[4]This is the difference between the premiums from the best fitting distribution and the subsidized actual premium.

the premiums are presented in Table 5.9. The mean absolute percentage differences between the subsidized premiums and the distribution based premiums are smaller (33%) than the difference between the unsubsidized premiums and the distribution based premiums (58%). Furthermore, the pairwise Wilcoxon signed-rank test shows that the subsidized premiums and the premiums based on the best-fitting distribution are not significantly different at 1% level.

Table 5.9. Mean Absolute-Percentage Difference in Actuarially Fair Premiums: Subsidized, Unsubsidized and Distribution Based

	Unsubsidized	Subsidized	Best-Fitting
Unsubsidized	-	(51.29)%***	(58.31)%***
Subsidized		-	32.92%**
Best-Fitting			-

Note: *** indicates significant at a 1%; **indicates significant at a 5% level under Wilcoxon signed ranks test; and a parenthesis indicates a negative difference.

Before ending this section, it is important to note that the procedures for calculating the distribution-based AFPs is fundamentally different from the approach used to calculate the actual premiums charged by RMA. For example, RMA rate-making procedures rely heavily on county level loss experience data and it individualizes rates based on the producer's historical yield. Actuarially, the use of actual loss experience data is a good attribute of the current rate making process and could be argued to contribute to the accuracy of this type of rating procedure. On the other hand, our distribution-based AFPs do not consider actual loss experience. Hence, it is difficult to ascertain which premium rating procedure truly represents individual loss risk the best. It may not necessarily be the case that the distribution-based AFP is the best procedure for setting premiums for existing crops (like corn), especially if there are actual loss experience data available for rating. But for the case of new crops to be insured, the rating procedure based on the Johnson family of distribution may be an excellent approach for developing premium rates for new insurance products designed for this crop. New crop insurance products typically do not have adequate loss experience data that can be used for accurate rating. Thus, as long as there is a reasonable amount of yield data for these new crops, setting premiums based on the Johnson family of distribution would be appropriate. This is especially important at this time because there is pressure from the government to expand the crop insurance program to include new crops, new regions, new products (such as livestock and timber) and new coverage (such as bioterrorism risks) (Glauber, 2004).

CHAPTER VI

SUMMARY AND CONCLUSIONS

Several alternative yield distributions have been offered by agricultural economists for use in risk related farm management research. Specifically, a number of nonparametric and parametric distributions have been proposed in the literature as a way to model crop yield behavior. Although nonparametric approaches have gained support in recent years due to its flexibility, the common problem of limited farm data have hindered the wider use of the nonparametric model in empirical work. In light of this, Ramirez and McDonald (2005) proposed the use of the Johnson family of distribution as a highly flexible parametric alternative to model crop yield behavior. The Johnson family of distributions includes the lognormal, S_U and S_B distributions. Theoretically, the S_U accounts for a wide-range of skewness-kurtosis combinations above the lognormal line, while S_B accounts for skewness-kurtosis combinations below the lognormal line. This enables the Johnson family of distribution to accommodate a wider range of possible skewness-kurtosis combinations. This study seeks to provide insight on the statistical performance of the Johnson family of distributions in terms of modeling crop yields from Illinois corn farms. In addition, this study examines the crop insurance premium rate implications of using the Johnson family of distributions. The approach used in this study is the comparison of the statistical performance and premium rate impacts of using the Johnson family of distributions to model yields versus the case where the more commonly used beta distribution is utilized. In addition, the premium rates calculated using the Johnson family of distributions was also compared with the unsubsidized/subsidized actual premium rates currently used in the crop insurance program.

The evaluation procedure requires that the candidate distributions considered be re-parameterized so that the skewness and kurtosis parameters are independent of the mean and the variance. This method permits the model to have mean and variance as a function of time and to allow inter-temporal shifts in both skewness and kurtosis. In so doing, all the yield distributions considered can address potential heteroskedasticity problems and allow for valid comparisons among the models. Maximum likelihood estimation was used to estimate the parameters of the candidate distributions and comparison of likelihood values was utilized to assess the statistical performance of the alternative models. The approach simply required choosing the model with the highest likelihood value. To demonstrate the economic impact of using alternative parametric distributions to model yields, a simulation model was then used to determine the impacts of using the Johnson family of distribution in setting premium rates for the Actual Production History (APH) insurance program.

The results of the study illustrated that the Johnson family of distributions may indeed be a highly flexible parametric approach to accurately estimate crop yield distributions. Using high quality data from Illinois corn farms, the S_B and S_U distributions statistically outperform the beta distribution with regard to modeling farm level yields (on average). The results also showed that using the more flexible Johnson family of distributions to model yield leads to a crop insurance premium rate that is significantly different from a premium rate established using the beta distribution. Actual unsubsidized premium rates using the RMA rating method also tend to be significantly different from the premium rates established using the Johnson family of distributions. This may be suggestive of adverse selection in the crop insurance program. However, when the premium rates based on the Johnson family was compared to the subsidized actual premium rates, the magnitude of the difference in the premiums became smaller. The subsidies implemented by the government to encourage participation of low-risk producers seemed to have the positive side-effect of reducing adverse selection in the program.

Limitations of the Study and Suggestions for Further Research

Although our research findings have important implications for modeling crop yields and crop insurance rating, many important research issues remain. Most fundamentally, our study focused only on comparing the Johnson family of distributions with only the beta distribution (which is considered the most common parametric distribution used to model yields). However, it would also be insightful to compare the Johnson family of distributions with other commonly cited parametric distributions used to model crop yields (i.e. Weibull, Gamma, Burr). For example, in Sherrick et al. (2004), the Weibull distribution was found to outperform the beta distribution for the same set of Illinois farms investigated in this study and, thus, the performance of the Weibull distribution relative to the Johnson family would be informative.

While the heteroskedasticity problem was accounted for in the modeling procedures used in the study, the autocorrelation problem was not explicitly considered (although the mean and variance of the model were modeled to be functions of time). Thus, for future studies, it may be useful to develop an explicit approach to address potential autocorrelation problems when using the Johnson family of distributions. Finally, the research analyses in this study were limited to modeling yield distributions at the farm level and for only one crop (corn). Although the data utilized captures a variation of yield behavior, it may be useful to further examine the performance of the Johnson family in modeling yields of other crops (such as soybeans, wheat, and specialty crops) or other crop insurance products (such as Crop Revenue Coverage, Revenue Assurance). Note, however, that assessing the revenue-based crop insurance program would require the estimation of joint price-yield distributions.

REFERENCES

Agricultural Outlook. "Risk Management" *Economic Research Service*, USDA, August 1999.

Babcock, B.A., C.E. Hart and D.J. Hayes. "Actuarial Fairness of Crop Insurance Rates with Constant Rate Relativities." *American Journal of Agricultural Economics* 86(August 2004): 563-575.

Balakrishnan, N. and V.B. Nevzorov. *A Primer on Statistical Distributions.* John Wiley & Sons, Inc., Publication, 2003.

Botts, R.R. and J.N. Boles. "Use of Normal-Curve Theory in Crop Insurance Rate Making." *Journal of Farm Economics* 40(May 1958):733-40.

Bradley, J. V. 1968. *Distribution-Free Statistical Test.* Prentice-Hall, Inc.

Chambers, R.G. "Insurability and Moral Hazards in Agricultural Insurance Markets." *American Journal of Agricultural Economics* 71(August 1989):604-616.

Chen, S. and M. Miranda. "Modeling Multivariate Crop Yield Densities with Frequent Extreme Values." Paper presented at the *American Agricultural Economics Association Annual Meeting*, Denver, Colorado, August 1-4, 2004.

Coble, K.H., T.O. Knight, R.D. Pope, and J.R. Williams. "An Expected Indemnity Approach to the Measurement of Moral Hazard in Crop Insurance." *American Journal of Agricultural Economics* 79(February 1997):216-26.

____________ . "Modeling Farm-Level Crop Insurance Demand with Panel Data." *American Journal of Agricultural Economics* 78(May 1996):439-447.

Day, R. H. "Probability Distributions of Field Crops." *Journal of Farm Economics* 47(August 1965):713-41.

Field, J.F. "Evaluating Crop and Revenue Insurance Products as Risk Management Tools for Texas Cotton Producer." Unpublished manuscript, Texas Tech University, Lubbock, TX, 2000.

Featherstone, A.M. and T.L. Kastens. "Non-Parametric and Semi-Parametric Techniques for Modeling and Simulating Correlated, Non-Normal Price and Yield Distributions: Applications to Risk Analysis in Kansas Agriculture." *Journal of Agricultural and Applied Economics* 32(August 2000):267-281.

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

Glauber, J. W. "Crop Insurance Reconsidered." *American Journal of Agricultural Economics* 86(2004):1179-1195.

Goodwin, B.K. "Premium Rate Determination in the Federal Crop Insurance Program: What Do Averages Have to Say About Risk?" *Journal of Agricultural and Resource Economics* 19(1994):382-95.

____________ . "Pricing Crop Insurance." Paper presented, September 13, 2002. Online. Available http://www.arec.umd.edu/policycenter/Crop-Insurance-and-Farm-Policy-Conference/presentations/goodwin.ppt#256,1,Pricing Crop Insurance (assessed April 2005).

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

Goodwin, B.K and O. Mahul. "Risk Modeling Concepts Relating to the Design and Rating of Agricultural Insurance Contracts." *World Bank Policy Research Working Paper* 3392, September 2004.

Goodwin, B.K., M.C. Roberts and K.H. Coble. "Measurement of Price Risk in Revenue Insurance: Implications of Distributional Assumptions." *Journal of Agricultural and Resource Economics* 25(July 2000):195-214.

Hoaglin, D., F. Mosteller and J. Tukey. *Exploring Data Tables, Trends and Shapes*. New York: John Wiley & Sons, Inc., 1985.

Insurance Information Institute. Online. Available http://www.iii.org/media/hottopics /insurance/crop/ (assessed April 2005).

Islam, Z., C. Turvey, and M. Hoy. "A Model of Agricultural Insurance in Evaluating Asymmetric Information Problems." Working Paper Series, Dept. of Agricultural Economics and Business, University of Guelph, Ontario, Canada. January 1999.

Johnson, N.L. "Systems of Frequency Curves Generated by Method of Translation." *Biometrika* 36(1949):149-176.

Johnson, N.L. and S. Kotz. *Continuous Univariate Distributions -1*. John Wiley & Sons, Inc. 1970.

Josephson, G.R., R.B. Lord, and C.W. Mitchell. "Actuarial Documentation of Multiple Peril Crop Insurance Ratemaking Procedures." Report prepared for the U.S. Department of

Agriculture, Risk Management Agency, by Milliman and Robertson, Inc. 2000. Online. Available http://www.rma.usda.gov/pubs/2000/MPCI_Ratemaking.pdf (assessed May 2005).

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

Just, R.E., L. Calvin, and J. Quiggin. "Adverse Selection in Crop Insurance: Actuarial and Asymmetric Information Incentives." *American Journal of Agricultural Economics* 81(1999):834-849.

Ker, A. P. "Private Insurance Company Involvement in the U.S. Crop Insurance Program." *Canadian Journal of Agricultural Economics.* 49(2001):557-566.

Ker, A.P. and B.K. Goodwin. "Nonparametric Estimation of Crop Insurance Rates Revisited." *American Journal of Agricultural Economics* 83(May 2000):463-478.

Ker, A.P. and K. Coble. "On the Use of the Beta for Modeling Yields." Working Paper, May 1998.

Knight, T.O. and K.H. Coble. "A Survey of Multiple Peril Crop Insurance Literature Since 1980." *Review of Agricultural and Applied Economics* 19(1997):128-56.

Makki, S.S. and A.L. Somwaru. "Asymmetric Information in Cotton Insurance Markets: Evidence from Texas." Paper presented at the *American Agricultural Economics Association Meetings*, Long Beach, California, July 27-31, 2002.

____________ . "Asymmetric Information in the Market for Yield and Revenue Insurance Products." Economic Research Service, U.S. Department of Agriculture. Technical Bulletin No. 1892 (April 2001).

Miller, S.E., K.H. Kahl, and P.J. Rathwell. "Evaluation of Crop Insurance Premium Rates for Georgia and South Carolina Peaches." *Journal of Agribusiness*. 18,3(Fall 2000a):303-317.

____________ . "Revenue Insurance for Georgia and South Carolina Peaches." *J. of Ag. and Applied Econ.* 32,1(April 2000b):123-132.

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

Norwood, B., M. Roberts and J. Lusk. "How are Crop Yields Distributed?" Paper presented at the *American Agricultural Economics Association Meetings*, Long Beach, California, July 27-31, 2002.

____________ . "Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions." *American Journal of Agricultural Economics* 86(November 2004): 1032-1043.

Pope, R.D. and R.F. Ziemer. "Stochastic Efficiency, Normality and Sampling in Agricultural Risk Analysis." *American Journal of Agricultural Economics* 66(February 1984): 31-40.

Ramirez, O.A. "Estimation and Use of a Multivariate Parametric Model for Simulating Heteroscedastic, Correlated, Non-Normal Random Variables: The Case of Corn-belt Corn, Soybeans and Wheat Yields." *American Journal of Agricultural Economics* 79(February 1997):191-205.

____________ . "Parametric Modeling and Simulation of Joint Price-Production Distributions under Non-Normality, Autocorrelation and Heteroscedasticity: A Tool for Assessing Risk in Agriculture." *Southern Agricultural Economics Association* 32(August 2000):283-297.

Ramirez, O.A. and T. McDonald. "The Reparameterized Johnson System: A Highly Flexible Yield and Price Distribution Model." *Department of Agricultural Economics and Agricultural Business and Extension Economics New Mexico State University Working Paper*, 2005.

____________ . "Ranking Crop Yield Models: A Comment." forthcoming.

Ramirez, O.A., C.B. Moss, and W.G. Boggess. "Estimation and Use of the Inverse Hyperbolic Sine Transformation to Model Non-Normal Correlated Random Variables." *Journal of Applied Statistics* 21:4 (1994):289-305.

Ramirez, O.A., S.K. Misra, and J.E. Field. "Crop yield Distributions Revisited." *American Journal of Agricultural Economics* 85(February 2003):108-120.

Risk Management Agency. Statement by R.J. Davidson, Jr. Administrator before the committee of an Agriculture, Rural Department, Food and Drug Administration and Related Agencies. 2005a.

____________ . "Crop Insurance Program Models." Online. Available http://www.rma.usda.gov/pilots/miscproj/cropmodl2.pdf. (assessed April 2005b).

Rodriguez, R.N. "A guide to the Burr Type XII Distributions." *Biometrika* 64(April 1977): 129-134.

Rothschild, M. and J. Stiglitz. "Equilibrium in Competitive Insurance Markets: an Essay on the Economics of Imperfect Information." *American Journal of Agricultural Economics* 90(November 1976):629-649.

Sherrick B.J., F.C. Zanini, G.D. Schnitkey, and S.H Irwin. "Crop Insurance Valuation Under Alternative Yield Distributions." *American Journal of Agricultural Economics* 86(May 2004):406-419.

Skees, J.R. and M.R. Reed. "Rate Making for Farm-Level Crop Insurance: Implications for Adverse Selection." *American Journal of Agricultural Economics* 68(August 1986):653-659.

Tadikamalla, P.R. and Johnson, N.L. "Systems of Frequency Curves Generated by Transformations of Logistics Variables." *Biometrika* 69(1989):461-465.

Tirupattur, V., R.J. Hauser and N.M. Chaherli. "Crop Yield and Price Distributional Effects on Revenue Hedging." *Office of Futures and Options Research*, 5(1996):1-17.

Turvey C.G. and J. Zhao. "Parametric and Non-Parametric Crop Yield Distributions and their Effects on All-Risk Crop Insurance Premiums." Department of Agricultural Economics and Business at the Univeristy of Guelph, Ontario, January 1999.

Wooldridge, Jeffrey M. *Introductory Econometrics: A Modern Approach*. 2nd ed. Ohio: South-Western Thomson Learning, 2003.

APPENDIX

Table A..1. Actual Premium Rate at 85% Coverage Level and 100% Price Election, $ per unit of liability

Farm	County	Lognormal	S_U	S_B	Beta	Actual Premium
a	De Kalb	0.036	0.052	0.043	0.023	0.088
b	De Kalb	0.034	0.044	0.049	0.033	0.130
c	De Kalb	0.031	0.048	0.034	0.032	0.104
d	La Salle	0.040	0.057	0.055	0.050	0.090
e	Wabash	0.075	0.099	0.094	0.066	0.152
f	De Witt	0.061	0.081	0.119	0.093	0.082
g	De Witt	0.012	0.036	0.044	0.031	0.076
h	Macon	0.052	0.036	0.045	0.067	0.084
i	La Salle	0.028	0.042	0.037	0.036	0.098
j	Champaign	0.023	0.059	0.039	0.069	0.087
k	Champaign	0.067	0.129	0.120	0.066	0.093
l	Champaign	0.038	0.047	0.047	0.051	0.095
m	Champaign	0.051	0.068	0.064	0.065	0.087
n	Douglas, Moultrie	0.041	0.043	0.052	0.042	0.098
o	Piatt	0.059	0.100	0.093	0.098	0.083
p	Piatt	0.084	0.105	0.105	0.101	0.088
q	Piatt	0.024	0.050	0.071	0.037	0.081
r	Piatt	0.056	0.080	0.071	0.072	0.094
s	Piatt	0.052	0.067	0.062	0.062	0.085
t	Moultrie	0.045	0.085	0.046	0.046	0.084
u	Vermilion	0.108	0.137	0.139	0.137	0.129
v	Sangamon	0.023	0.062	0.076	0.054	0.085
w	Sangamon	0.017	0.023	0.043	0.048	0.081
x	Menard	0.146	0.138	0.130	0.130	0.112
y	Sangamon, Mccoupin	0.047	0.072	0.062	0.068	0.084
z	Vermilion	0.076	0.111	0.046	0.132	0.144
Average		0.051	0.072	0.069	0.066	0.097
Std. Dev.		0.030	0.033	0.031	0.032	0.020
Minimum		0.012	0.023	0.034	0.023	0.076
Maximum		0.146	0.138	0.139	0.137	0.152

PERMISSION TO COPY

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Texas Tech University or Texas Tech University Health Sciences Center, I agree that the Library and my major department shall make it freely available for research purposes. Permission to copy this thesis for scholarly purposes may be granted by the Director of the Library or my major professor. It is understood that any copying or publication of this thesis for financial gain shall not be allowed without my further written permission and that any user may be liable for copyright infringement.

Agree (Permission is granted.)

Catherine P. Lawas 11/20/05

Student Signature Date

Disagree (Permission is not granted.)

Student Signature Date