SEMINONPARAMETRIC ESTIMATION

OF CONDITIONAL YIELD DENSITIES

by

Hao Lu

_____

A Thesis Submitted to the Faculty of the

DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2009

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona.

SIGNED: Hao Lu

APPROVED BY THE THESIS DIRECTOR

This thesis has been approved on the date shown below:

Alan P. Ker

Professor

Dept. Agricultural and Resource Economics

07/10/09

Date

ACKNOWLEDGEMENTS

Dr. Alan Ker, my advisor, deserves many thanks for his constructive guidance and unreserved help throughout this study. Similarly, I am thankful to my committee members, Dr. Satheesh Aradhyula and Dr. Tauhidur Rahman, for spending their precious time helping me improve this thesis.

I would also like to thank other professors in the Department of Agricultural and Resource Economics who have taught me during the past two years: Dr. Bruce Beattie, Dr. Roger Dahlgran, Dr. Paul Wilson, Dr. Chris Stefanadis, and Dr. Gary D. Thompson. Their instructions are inevitable in the growth of my knowledge and skills. I am also thankful to the graduate coordinator Nancy Smith. She has provided countless helps since the first time I came to the department.

Finally, I would like to dedicate this thesis to my parents, Zhao Xinhua and Lu Yinglin.

Tucson, Arizona, USA Hao Lu

July 2009

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Government sponsored agricultural insurance programs are susceptible to losses due to moral hazard and adverse selection problems. To lessen the problems, the Group Risk Plan (GRP) of insurance has been offered by the Federal Crop Insurance Corporation since 1993. Under this product, losses are measured on the basis of a county's mean yield. Accurate pricing of insurance products like GRP requires the distribution of future yields conditional on the current information set. The main objective of this thesis is to improve the accuracy of premium rates thus improving the efficiency of GRP program. In satisfying this objective, a seminonparametric (SNP) maximum likelihood method is utilized in an attempt to reduce program inefficiencies induced by distributional assumptions in determining premium rates. A spline model with one knot point is used to capture the central tendency and to predict mean yields. A mixture of two normal distributions is used to represent the disturbance distribution.

National Agricultural Statistics Service (NASS) county mean yield data over the period 1955 to 2007 of 102 Illinois counties are used in the analysis. The crop analyzed is corn for grain. The premium rates are calculated numerically from the recovered yield distributions of each county. The SNP maximum likelihood approach makes more efficient use of the data in that yield correlation among counties is explicitly modeled. As a result, SNP rates tend to be more consistent among counties.

CHAPTER ONE

INTRODUCTION

Government sponsored agricultural insurance programs are susceptible to losses due to moral hazard and adverse selection problems. Multiple Peril Crop Insurance (MPCI), the traditional insurance product of the Federal Crop Insurance Corporation (FCIC), has been based on individual producer yields. To lessen the problem of moral hazard and adverse selection, Group Risk Plan (GRP) has been offered by the FCIC since 1993. Under this product, losses are measured on the basis of a county's mean yield. That is, producers may only purchase insurance against a realization of their county's mean yield below the predicted mean yield. GRP is not offered in those counties where individual producers can influence the county mean yield. Thus, moral hazard problem is eliminated under GRP. Adverse selection may be reduced because information on county yield is generally available and more reliable than that on individual yield. However, misuse of county yield data in constructing GRP premium rates may not reduce losses but simply transfer adverse selection responses at the farm-level to the county-level.

Accurate pricing of insurance products under either MPCI or GRP requires the distribution of future yields conditional on the current information set. Proper representation of the conditional yield distribution may be complicated by several factors affecting both the location and the shape of the distribution. For instance, biological constraints limit the maximum yield attainable while environmental factors tend to adversely affect yield levels. Hence, crop yields distributions tend to be negatively skewed. In addition, local idiosyncrasies may affect not only the location and scale of the distribution but the entire shape of the distribution. Hence, the set of county specific

conditional distributions may not belong to a common distributional family. If the distributional assumption is incorrect, probability estimates are biased, and hence premium rates are incorrect. As a result, low-risk producers who are overcharged are less likely to purchase insurance, while high-risk producers who are undercharged are more likely to over-purchase insurance.

A secondary policy problem is maintaining consistency of premium rates among contiguous counties. There should be consistency to some extent since yield realizations would be similar. However, given the limited number of realizations, large yield deviations among contiguous counties may lead to significantly different rates. Properly incorporating contiguous county information into the construction of premium rates is another problem addressed in this thesis.

The main objective of this thesis is to improve the accuracy of premium rates thus improving the efficiency of GRP program. As stated above, accurate rates require proper representation and estimation of the conditional yield densities. In satisfying this objective, seminonparametric (SNP) maximum likelihood method is utilized in an attempt to remove any program inefficiencies induced by distributional assumptions in determining premium rates. A spline model with one knot point is used to capture the central tendency and to predict mean yields. SNP method estimates both the spline model and a separate disturbance distribution for each county simultaneously. A mixture of two normal distributions is used to represent the disturbance distribution. This structure can accommodate various departures from normality that exist in the conditional yield distributions (i.e. skewness, multi-modality). Parameters indexing the mixture distribution follow a mixed effects regression model where the joint density of the

random effects is estimated by a Hermite expansion. County specific yield distributions are completely defined given the empirical Bayes estimates of the random effects and estimates of other fixed effects. At the same time, correlation among counties is explicitly modeled.

The remainder of the thesis has four chapters. Chapter two reviews insurance in U.S. agriculture, in particular GRP. Chapter three provides a review of the literature on modeling conditional yield densities. Chapter four describes the data used in the empirical analysis, specifies the structure of the model, and outlines seminonparametric maximum likelihood estimation techniques. The last chapter reports the empirical results, summarizes the findings, and discusses areas requiring further research.

CHAPTER TWO

INSURANCE IN AGRICULTURE

2.1 Insurance in U.S. Agriculture

Insurance is an instrument to spread risk over time, across space, and among different industries and individuals. Farming has always been an inherently risky industry because it depends on natural conditions and is subjected to various weather related perils such as droughts, flood, hurricanes and other natural disasters. Since the 1930s, many American farmers have been able to transfer part of the risk of loss in production to the federal government through crop insurance.

Congress first authorized Federal crop insurance in the 1930s along with other initiatives to help agriculture recover from the combined effects of the Great Depression and the Dust Bowl. The Federal Crop Insurance Corporation (FCIC) was created in 1938 to carry out the program. Initially, the program was started as an experiment, and crop insurance activities were mostly limited to major crops in the main producing areas. Crop insurance remained an experiment until passage of the Federal Crop Insurance Act of 1980 (RMA, USDA).

Coble and Knight (2002) describe the Federal Crop Insurance Act of 1980 as the foundation of the modern era of U.S. crop insurance. The act helped start the expansion of the program to include many more crops and regions, and most importantly, more farmers' participation by authorizing premium subsidies and allowing private sector delivery of federal crop insurance. However, during the 1980s and up through 1993, the crop insurance program witnessed a consistent high loss ratio (indemnities/total premiums; as high as 2.5 in the drought year of 1988) and failed to prevent ad hoc

disaster assistance in several years. Therefore, the Federal Crop Insurance Reform Act was enacted in 1994. It created catastrophic (CAT) coverage, a low-coverage and effectively free insurance and increased subsidies for higher coverage level insurance. Farmers were required to purchase crop insurance or otherwise waive their eligibility for any disaster benefits that might be available. Following the 1994 Act, participation in the crop insurance program increased significantly. According to RMA USDA, the acreage of farmland under insurance in 1998 is more than three times that in 1988, and more than twice that in 1993.

Most recently in 2000, the Agricultural Risk Protection Act provided the private sector an increased role in developing new crop insurance products and further increased the premium subsidies to encourage producers to purchase higher insurance coverage levels and to make the insurance program more attractive to prospective producers. Figure 1 summarizes some attributes of the U.S. crop insurance program including the loss ratio over the 1999-2008 period and premium combination during the 2004-2008 period. The program is expected to provide about $71.8 billion in risk protection on about 288 million acres in 2009. This represents about 80 percent of the nation's acres planted to principle crops. The 2009 budget of the total government cost of the crop insurance program achieves $6.4 billion (USDA, 2008).

Given the importance of crop insurance to U.S. agricultural sector and the increasing government resources devoted to crop insurance, there are more concerns about the cost efficiency of the program and its vulnerability to fraud, waste, and abuse. As with most insurance, crop insurance is susceptible to moral hazard and adverse selection problem.

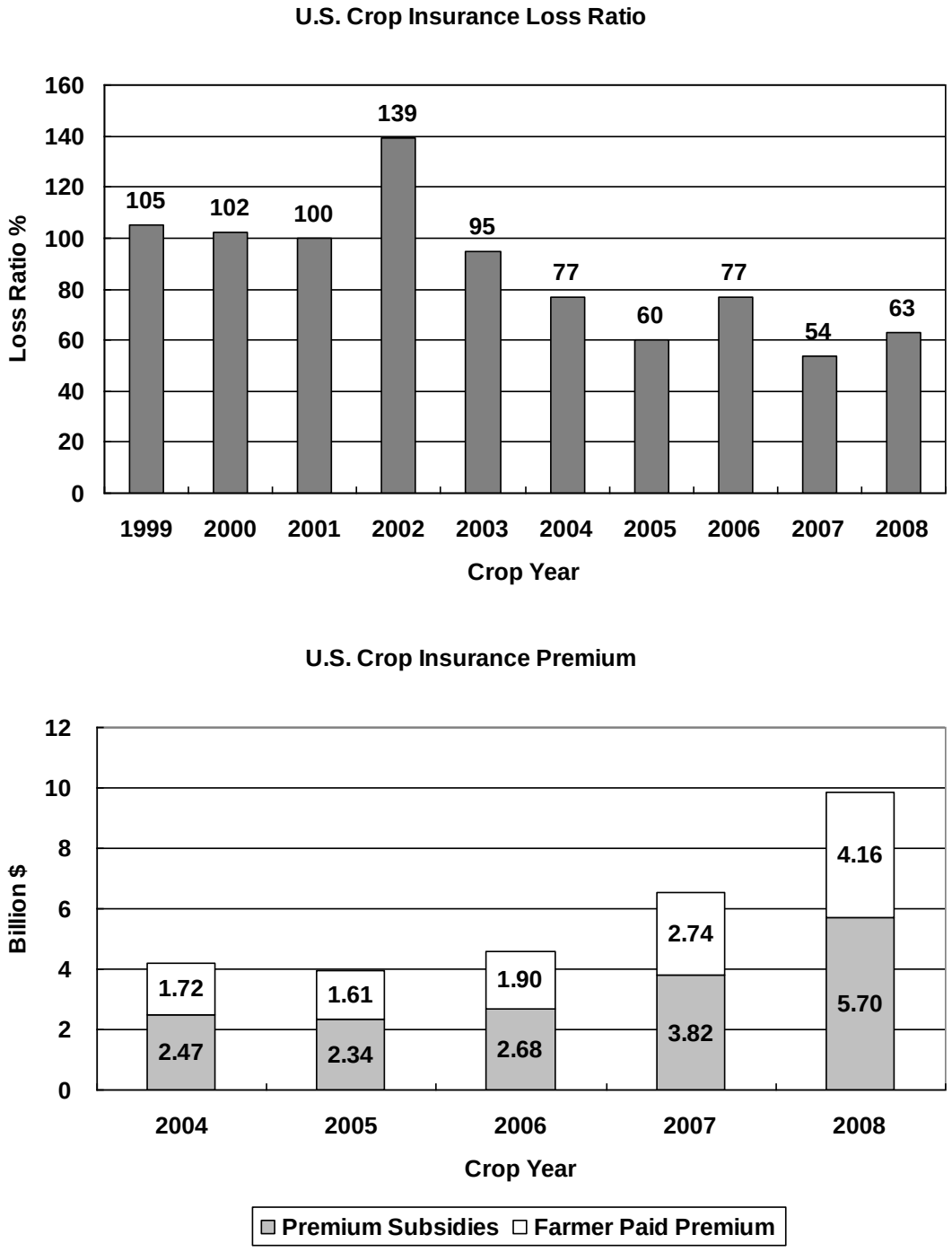


Figure 1: Summary of U.S. Crop Insurance Loss Ratio 1999-2008 and Premium 2004-2008

Source: USDA Risk Management Agency

2.2 Moral Hazard and Adverse Selection

Moral hazard occurs when an insured producer can increase his or her expected indemnity by actions taken after buying insurance (Glauber and Collins, 2002). FCIC can hardly monitor production practices throughout the growing season. A farmer may fail to fertilize properly or treat for pests. Therefore, the farmer actually changes his yield distribution by actions unknown to FCIC and may benefit from it. U.S. GAO (2007) has reported some farmers may have abused the crop insurance program by allowing crops to fail through neglect or deliberate actions in order to collect insurance.

Adverse selection occurs when an insured producer has more information about the risk of loss and thus is superior at estimating the fair premium rate than the insurer. An actuarially fair premium rate is a rate that is set such that premiums collected are equal to expected indemnities. By and large, farmers know more about their yield distributions than insurers. As a result, low-risk producers who are overcharged are less likely to purchase insurance, while high-risk producers who are undercharged are more likely to over-purchase insurance. Over time, this will distort participation in favor of the higher risks, and thus premiums will not be sufficient to cover indemnity payments. And raising premium for all producers will create an even more adversely selected market since less risky participants drop out of the program. Deng, Barnett, and Vedenov (2007) point out that the premium rates of farm yield insurance often contain large positive wedges[1] caused by moral hazard and adverse selection. The large positive wedge (negative expected value) will make the insurance product unattractive to many potential insureds.

[1] The term wedge is used to describe the difference between the premium cost and the expected indemnity for an insurance product. A positive wedge implies that the premium cost is greater than the expected indemnity.

2.3 Crop Insurance Products

A crop insurance contract is a commitment between insured farmers and their insurance providers. Under the contract, the insured farmer agrees to insure all the eligible acreage of a crop planted in a particular county. This choice is made county by county and crop by crop. All eligible acreage must be insured to reduce the potential for adverse selection against the insurance provider. The insurance provider agrees to indemnify (or protect) the insured farmer against losses that occur during the crop year. Losses must be due to unavoidable perils beyond the farmer's control.

A variety of crop insurance products are currently available to farmers and a number of new pilot programs are under development. Multiple Peril Crop Insurance (MPCI) is the traditional product provided by FCIC. This program insures producers against yield losses due to natural causes such as drought, excessive moisture, wind, insects and disease. Coverage is available at 50% to 75% of the predicted yield for the farm. The predicted yield is determined from producer production records for a minimum of 4, up to 10 consecutive crop years. For producers who provide less than 4 years of actual yields, a T yield is substituted for each missing year. [2] After 10 years of history are reached, the predicted yield becomes a moving 10-year average yield. The farmer also selects the percent of the predicted price he or she wants to insure; between 55 and 100 percent of the FCIC expected market price. If the harvest is less than the yield insured, the farmer is paid an indemnity based on the difference. Indemnities are calculated by multiplying this difference by the selected price level.

[2] Each county has a different T yield. It is based on the 10-year historical county average yield. Growers with no records are assigned 65% of the T yield as their predicted yield. Growers with a record for one year, two years, and three years receive 80%, 90%, 100% of the T yield respectively for the missing years. Once each year has been assigned a yield, the predicted yield is just a simple average of the four yields.

Optimally, an insurance provider would prefer to calculate individual premium rate for each farmer on the basis of that farmer's risks and expected yields. However, individual data are rare at best. Information asymmetry about individual yield distributions leads to adverse selection responses while information asymmetry about producers' actions exposes MPCI to moral hazard problems. One way to lessen the problem is to base the crop insurance products on more aggregate data, such as data at the county level. In fact, such index-based crop insurance products were developed to overcome the problem of short or nonexistent individual crop yield data, which represented a dramatic departure from conventional crop insurance products.

Since 1993 an area yield insurance product called Group Risk Plan (GRP) has been offered through the FCIC for selected crops and regions (Deng, Barnett, and Vedenov, 2007). GRP is designed as a risk management tool to insure against widespread loss of production of the insured crop in a county. The product was developed on the basis that if producer yields are highly correlated within the given area, area-yield insurance covers approximately the same risk as farm-level insurance. That is, when an entire county's crop yield is low, most farmers in that county will also have low yields and thus, farmers would receive an indemnity payment. GRP uses a county index as the basis for determining a loss. When the county yield for the insured crop, as determined by National Agricultural Statistics Service (NASS), falls below the trigger level chosen by the farmer, an indemnity is paid. In this sense, GRP is essentially a put option on the average yield for a production region. Producers must choose one coverage level for each crop and county combination. The grower selects the dollar amount of protection per acre (between 90% and 150% of the forecasted value) and one of the five coverage levels (70,

75, 80, 85, or 90%) of the FCIC expected county yield. The expected county yield is calculated using many years of county data from the NASS with an adjustment for the yield trend.

GRP has the potential to reduce losses from moral hazard and adverse selection problems. By definition, moral hazard is eliminated with GRP because actions of individual producers hardly exert noticeable influences on county yields. GRP is only offered in counties where this conjecture is reasonable. Adverse selection may be reduced because information on county yield distributions is generally available and more reliable than that on individual yield distributions. However, Ker (1996) points out that using this information to construct actuarially fair premium rates is difficult and misuse of county yield data in constructing GRP rates may not reduce losses but simply transfer adverse selection responses at the farm-level to the county-level. GRP also has the advantage of cost efficiency compared to farm-level insurance. Producers do not have to maintain and provide production history or evidence of loss because payments are made on losses based on the county yield. The only information a producer needs to provide is the number of acres planted.

Because farm-level yields are not perfectly correlated with the area average yield, GRP purchasers are exposed to basis risk (Deng, Barnett, and Vedenov, 2007). It is possible for a farmer to experience a production loss and yet not receive an indemnity because there has been no shortfall in the county average yield. In general, GRP basis risk is higher in regions with more heterogeneous production conditions, such as elevation, soil type, drainage, and climate. Skees, Black, and Barnett (1997) gives an example of such a region—the sub-mountainous deciduous fruit production region

around Hood River, Oregon. In a sub-mountainous region, freeze is the major source of yield risk, and the probability and extent of yield loss depend largely on localized topographical features such as elevation. Area yield insurance will not provide effective risk management in an area such as this.

2.4 Premium Rate

Accurate pricing of insurance products and precise risk assessment is critical to the success of achieving an actuarially sound crop insurance program. An actuarially fair premium rate is a rate that is set such that premiums collected are equal to expected indemnities. Indemnity is paid when the value of a farmer's production is less than the liability he or she purchased. The amount of indemnity paid is equal to the difference between the liability and the value of production (VP). For a farmer who selects a coverage level of λ percent of the FCIC expected yield y^e (bushels/acre), plants A acres of land, chooses a price level of P dollars per bushel, and realizes a yield of Y (bushels/acre),

(1) $\text{Liabilty} = A \times \lambda y^e \times P \quad \text{VP} = A \times Y \times P$

$$\textit{if} \quad Y < \lambda y^e, \quad \text{Indemnity} = \text{Liabilty - VP} = A \times (\lambda y^e - Y) \times P$$

(2) Premium Rate (%) =

$$\frac{\text{Expected} \quad \text{Indemnity}}{\text{Liabilty}} = \frac{\Pr(Y < \lambda y^e)[\lambda y^e - \text{E}(Y \mid y < \lambda y^e)]}{\lambda y^e}$$

where the expectation operator and probability measure are taken with respect to the conditional yield density $f_Y(y \mid \text{I}_t)$, and I_t is the information set known at time of rating. Note that the acres planted A and price level P in both the numerator and denominator of equation (2) are cancelled out. Accuracy in premium rates is entirely dependent on

accurate estimation of the conditional yield densities, especially the lower tails—below λy^e—of the yield density estimates.

To construct premium rates, the conditional distribution of yields in time t + 2 given information available until time t needs to be recovered.[3] In general, this is done by first estimating the central tendency[4] and predicting the mean of the conditional yield distribution. Then, a distributional assumption is made (usually normal distribution) and the fitted errors from estimating the central tendency are used to recover the parameters indexing the assumed distribution. The conditional yield distribution is recovered by a simple location transformation of the error term distribution.[5]

In 1996, the use of spline regression was employed by FCIC for estimating central tendency. Empirical premium rates, sum of all percentage shortfalls divided by the number of years, were calculated at various coverage levels in the GRP. The implied coefficient of variation of an assumed normal yield distribution, based on the empirical premium rate at 100% coverage level, was put into the Botts and Boles (1958) algorithm to develop parametric premium rates for each levels of coverage used in the GRP. At each coverage level, the parametric premium rate was compared to the empirical premium rate and the higher rate was used in GRP contract. The final premium rate for each county GRP contract was a weighted average of the premium rate for that county and the premium rate for each contiguous county using relative acreage as weights (Skees,

[3] This lag reflects the substantial amount of time required to derive accurate aggregate yield measures from farm-level surveys.

[4] The first GRP contracts estimated central tendency using a robust double exponential smoothing procedure. In 1995, central tendency was estimated for wheat GRP using ARIMA models developed by Ker and Goodwin. In 1996, the method for estimating central tendency changed to spline regression with one knot point.

[5] Location transformation assumes a model with additive error term (i.e. $y = g(x) + e$).

Black, and Barnett, 1997). [6]

CHAPTER THREE

LITERATURE REVIEW

Crop yields follow a spatio-temporal process, in the sense that, if we take the average over some spatial region (field, farm, or county), conditional on the underlying temporal process, we can recover the conditional mean yield density for that given space at a point in time (Ker and Goodwin, 2000). In most empirical work, the only information known at time t is the time index and previously realized yields.

Over years, economists have conducted much research on modeling conditional yield densities. This chapter provides a review of the literature on the subject. Corresponding to the spatio-temporal characteristic of crop yields, the chapter is separated into two main parts: the spatial aspect of yields and the temporal process of yields. In addition, the spatial aspect part is further divided into two sections: parametric approaches and nonparametric approaches.

3.1 Spatial Aspect of Yields

In empirical applications, most researchers opt to use parametric method where a parametric distribution is assumed and the parameters indexing the distribution are estimated with the detrended yield data. The literature is replete with candidate yield distributions, few of which can be excluded on theoretical grounds. This has led economists to seek empirical evidence as to which models are best. Along another vein, some researchers try to solve the problem with the alternative nonparametric methods.

[6] A reserve loading was implemented by dividing the final premium rate by 0.9 in order to build reserves.

3.1.1 Parametric Approaches

Crop yield distributions are generally recognized as skewed (nonnormal). Ramirez (1997) claims that (i) adverse weather conditions, (ii) biophysical and technological limitations on the maximum yields attainable during any given year cause the yield distributions to exhibit left-skewness. Gallagher (1987) states that yield cannot exceed the biological potential of the plant, yet it can approach zero under blight, early frost, or extreme heat. Declining positive and negative marginal returns to weather inputs is one set of factors that could produce negatively skewed distributions of yields. More recently, Hennessy (2009) formalizes the role of the weather: whenever the weather-conditioned mean yield has diminishing marginal product with respect to a weather-conditioning index, then there is a disposition toward negative yield skewness. This is because high marginal product in bad weather stretches out the yield distribution's left tail relative to that for weather.

In addition, Nelson (1990) points out that spatial dependence between individual losses means that the Central Limit Theorem (CLT) cannot be used as a basis for assuming that crop yields are normally distributed. Hennessy (2009) formally explains why CLT does not help in any way to narrow down the set of distributions appropriate for crop yield modeling.

The calculation of a fair insurance premium rate involves integration under the lower tail of the distribution function and skewness can significantly alter the probability mass in that tail. Thus, use of a normal distribution could produce significant errors if the true distribution is skewed. Goodwin and Ker (1998) mention that densities that are more symmetric than the population will result in premiums that tend to be larger than they

should be for negatively skewed distributions. Using farm-level yield data of Iowa corn producers, Nelson (1990) shows that the symmetry of the normal distribution can cause significant overestimates of a premium when the true yield distribution is negatively skewed.

The beta distribution has been used by many researchers to accommodate skewness of yield distributions; see for example, Nelson and Preckel (1989); Nelson (1990); Coble et al., (1996); Babcock and Hennessy (1996). Nelson and Preckel (1989) list three reasons why crop yield may be distributed as beta distribution: (i) Crop yields are known to fall in a range from 0 to some maximum possible value. The beta variate may be defined on an interval $(0, y^u)$ where y^u is a finite upper bound on the random variable. (ii) Crop yield distributions can be significantly skewed either to the right or to the left. The beta distribution has such flexibility. (iii) The beta distribution is well known and mathematically tractable. All of the moments of the distribution exist and are simple functions that are ratios of polynomials in the parameters of the distribution.

Some researchers have assumed crop yields follow other distributional families. Gallagher (1987) uses the gamma distribution because varying degrees of skewness and variance can be captured with relatively few parameters. A positively skewed gamma variate corresponds to a negatively skewed yield deviation (error term is subtracted from capacity). However, Norwood, Roberts, and Lusk (2004) point out that gamma is sensitive to the maximum yield value chosen. When conducting real forecasts, it is difficult to identify the maximum yield needed to implement the gamma model.

Taylor (1990) presents two operational approaches to empirically fit multivariate nonnormal joint probability density functions. In lieu of assuming a specific multivariate

density for empirical analysis, a cubic polynomial approximation of the cumulative distribution function constrained to the zero-one range by a hyperbolic tangent is used. Empirical estimates of the joint distribution for corn, soybeans, and wheat yields are presented to illustrate the method. Ramirez (1997) specifies three limitations inherent in Taylor's method, including (i) impossibility to assess the flexibility of the technique, (ii) sacrificed statistical efficiency, (iii) incompatible with heteroskedasticity.

Moss and Shonkwiler (1993) uses an inverse hyperbolic sine transformation to model yields. Ramirez (1997) proposes a modified inverse hyperbolic sine transformation which extends the original parameterization to a multivariate nonnormal density function that simultaneously accounts for skewness, kurtosis, heteroskedasticity, and correlation among the random variables of interest. The theoretical attributes of the modeling tool are discussed and exemplified by analyzing and simulating Corn Belt corn, soybean, and wheat yields.

Nonetheless, not all researchers accept the idea of the skewness of crop yields. Just and Weninger (1999) attempt to restore the support for the normal distribution. They point out deficiencies in the work of Nelson and Preckel (1989), Taylor (1990), Moss and Shonkwiler (1993), and Ramirez (1997). Then they generalize three methodological problems common in yield distribution analysis: (i) misspecification of the nonrandom components of yield distributions, (ii) misreporting of statistical significance, and (iii) use of aggregate time-series data to represent farm-level yield distributions. After proposing their own procedure, they present tests of normality using both detrended aggregate time-series yield data and farm-level yield data from 106 Kansas farms during 1973-87. The results contradict previous findings of nonnormality.

In response, Atwood, Shaik, and Watts (2003) use Monte Carlo procedures to demonstrate that Just and Weninger's suggested detrending procedures often tend to bias the analysis in a Type II direction i.e., in favor of failing to reject normality for nonnormal error distributions. The results of Monte Carlo simulations using several nonnormal distributions demonstrated that the powers of the standard and omnibus normality tests are often significantly reduced when trends are individually estimated for a short-term yield panel data set. An alternative detrending procedure is suggested that demonstrates higher power for the distributions examined. Normality is generally rejected when both Just and Weninger's and the alternative procedures are applied to a larger Kansas panel data set than used by Just and Weninger.

In the same year, Ramirez, Misra, and Field (2003) utilize an expanded, refined parameterization of Johnson S_U family of densities, arguing that it is flexible enough to alleviate the concerns of using competing distributional assumptions in applied research. The parameters determining the first four moments of and the correlations between the yield distributions are jointly estimated by maximum likelihood. This expanded S_U family is used to revisit the issue of whether aggregate Corn Belt corn, soybean, and wheat yield distributions are nonnormal. Using an expanded data set and addressing the procedural issues that have been raised in Just and Weninger (1999), they reaffirm Ramirez's (1997) findings that Corn Belt corn and soybean yields are nonnormally distributed and left skewed.

Goodwin and Ker (2002) also disagree with Just and Weninger (1999). By conducting a simulation where the underlying distribution is nonnormal, they illustrate that while the statistical test will tend to fail to reject the normal, the economic

consequences of using the normal to derive crop insurance rates can be disastrous. They emphasize that the inability to reject normality (or any parametric form) based on a relatively small sample should never be used as grounds to assume normality.

Sherrick et al. (2004) recover corn and soybean yield distributions under alternative distributional representations. Farm-level data set from the University of Illinois Endowment Farms containing same-site yield records from 1972 to 1999 is used. Even though the detrended yields display predominantly negative skewness and fatter tails than would be implied by normal distributions, logistic, normal, and lognormal distributions are also considered in addition to beta and Weibull distributions. Formal goodness-of-fit test is conducted to assess the appropriateness of each distribution in representing the sample data for each farm. Beta and Weibull rank first and second overall, followed by the logistic, then normal and finally lognormal. It is also demonstrated that distributional choice can have large impacts on insurance valuation and risk assessments.

3.1.2 Nonparametric Approaches

Ker and Goodwin (2000) express two concerns about the parametric approaches. First, the fact that some parametric families can accommodate negative skewness does not indicate crop yields can be adequately approximated by them. The unknown yield distribution may or may not be uniquely defined by its first three population moments. Theoretical distributions may be constructed such that their resulting premium rates at low coverage levels differ by an order of magnitude despite having identical first three moments. Second, these parametric families do not allow bimodality. Central Limit

Theorem for dependent processes suggests the possibility that yields at some aggregate level may be bimodal. Goodwin and Ker (1998) found evidence of bimodality, suggesting that high yields (near the capacity constraint) happen frequently while relatively low yields also happen fairly often. Yields between the extremes are less likely.

Given the deficiencies associated with parametric forms, some economists have used the alternative nonparametric density estimation approach; see for example, Goodwin and Ker (1998); Deng, Barnett, and Vedenov (2007); Nadolnyak, Vedenov, and Novak (2008); Ker and Goodwin (2000); Racine and Ker (2006).

Nonparametric density estimation techniques do not assume a particular functional form for the yield distributions but rather allow the data to select the most appropriate representation of the yield distributions. Most nonparametric density estimation applications utilize the kernel method of smoothing. Under the kernel approach, each observation is surrounded by a symmetric weighting function *K*. A bandwidth parameter determines the weight to assign to neighboring observations in constructing the density and thus corresponds to the amount of smoothing to be done. A larger bandwidth will smooth more and thus will result in a flatter, smoother density function while a small bandwidth will yield a rough and irregular density (Goodwin and Ker, 1998).

Goodwin and Ker (1998) apply nonparametric density estimation methods to model county-level crop yield distributions. Expected yields are forecast and premium rates are calculated for the 1995-96 GRP for barley and wheat. In most cases, existing rates are significantly smaller than the nonparametric rates, suggesting that existing rating procedures may understate the risk of loss. Ker and Goodwin (2000) improve the

nonparametric kernel methods by employing empirical Bayes techniques on the estimated values from the kernel density estimator. Simulations suggest that this empirical Bayes nonparametric estimator may provide very significant efficiency gains in estimating conditional yield densities. Furthermore, a variable smoothing approach is employed which significantly decreases the dependency of estimated tail probabilities on the specific location of the tail realizations.

Norwood, Roberts, and Lusk (2004) rank six yield densities based on their out-of-sample forecasting performance. The forecasting ability for each density is ranked according to its likelihood function value when observed at out-of-sample observations. All six models are mentioned in this chapter, including Gallagher (1987), Nelson and Preckel (1989), Moss and Shonkwiler (1993), Ramirez (1997), Goodwin and Ker (1998), and Just and Weninger (1999). The first contest utilizes Corn Belt corn, soybean, and wheat yields from 1950 to 1989. The second contest utilizes county level yields from 1962 to 1992. Univariate yield distributions are estimated for corn, wheat, and soybeans. In both contests, the nonparametric model offered by Goodwin and Ker (1998) best forecast county average yields.

The limitations of nonparametric approaches are also acknowledged. Other than the requirement to choose a kernel function and a bandwidth parameter, the most prominent one is the lack of efficiency. Goodwin and Ker (1998) recognize that although the kernel procedures provide a consistent estimate of any density, the rate of convergence to the true density is relatively slow. If one has prior knowledge of the true parametric form of the density, a parametric specification will provide more efficient estimates. Ker and Coble (2003) claim that nonparametric methods tend to be inefficient

relative to maximum likelihood methods when the assumed parametric model is correct. It is possible, perhaps likely, for very small samples such as those corresponding to farm-level yield data, that an incorrect parametric form, say Normal, is more efficient than the standard nonparametric kernel estimator.

Ker and Coble (2003) propose a semiparametric estimator that encapsulates the benefits of both parametric and nonparametric methods while mitigating their disadvantages. The semiparametric estimator begins with a parametric estimate and then corrects it nonparametrically based on the data. If the assumed parametric model is correct the semiparametric estimator attains the parametric rate of convergence while if not, it attains the nonparametric rate of convergence. Most importantly, if the assumed parametric model is sufficiently close, efficiency gains are realized in relation to the standard nonparametric kernel estimator. In the simulation, for samples above fifteen the semiparametric estimator with the normal distribution is preferred to the competing parametric and nonparametric estimators.

3.2 Temporal Process of Yields

The use of time series data to represent a point-in-time distribution requires that the time component, if any, be controlled in the data. Approaches for doing so include deterministic trend models and stochastic trend models. Deterministic trend models are based on time functions such that the random effects from year to year have no effect on trend, while stochastic trend models have trends that are permanently affected by previous shocks (Sherrick et al., 2004).

Among deterministic time trends, the simplest and most widely used one is linear

trend of time; see for example, Gallagher (1987); Taylor (1990); Coble et al., (1996); Ramirez (1997); Sherrick et al. (2004). Just and Weninger (1999) express their concerns about the use of linear trend. They argue that misspecification of the deterministic component of yields, represented by differences between the true and assumed trend specifications, causes non-stationarity of yield deviations and incorrect assessment of skewness and kurtosis. They show that substantial departures in skewness and kurtosis from normality are possible in either direction by making a simple error in choosing the degree of polynomial. Instead of linear trends, they reveal quadratic or cubic polynomial trends when analyzing Ramirez's data and Gallagher's data.

Following the procedures outlined in Just and Weninger (1999), Sherrick et al. (2004) fit polynomial time trends of fifth order and lower and selecting the order based on the significance of *F*-tests for higher order terms. The result is that linear trends are imposed in the end. Similarly, to address the concerns raised in Just and Weninger (1999), Ramirez, Misra, and Field (2003) specify the means of the yield distributions as fourth- and third-degree polynomial functions of time for the Corn Belt and Texas Plains county-level yield data respectively. The final models include quadratic, linear, and cubic polynomial trends for the means of the corn, soybean, and wheat yield distributions, after excluding parameters not significantly different from zero.

Log-linear trends are utilized in Deng, Barnett, and Vedenov (2007), and Nadolnyak, Vedenov, and Novak (2008). Following the practice of FCIC, Ker and Coble (2003), Racine and Ker (2006) employ the one-knot linear spline to approximate the temporal process of yields. The justification is also provided: as new technologies become available, early adopters capture profits. As early adopters demonstrate to others

that profits are possible with the new technologies, a larger contingent adopts thus reducing profits. Finally, the remaining non-adopters either adopt or exit. The unused resources from those exiting are acquired by those remaining. Given that adopters have higher yields than those that exited, average yields must increase. The greater the rate of consolidation of these resources, the greater will be the rate of adoption. Thus, while technology will directly increase average yields, the rate of farm consolidation will also increase average yields. Farm consolidation was greatest in the decades of the 1950s and 1960s and slow in the 1970s thus suggesting a one-knot spline.

Moss and Shonkwiler (1993) propose modeling the central tendency of crop yield distribution with a stochastic trend model, which explicitly recognizes that the distribution's central moment need not evolve at a constant rate over time. The stochastic trend model is based on two equations, a measurement equation and a transition equation. The measurement equation relates changes in an observable variable (crop yield), to a vector of unobservable state variables, deterministic effects, and an error term. The evolution of the state vector over time is governed by the transition equation which includes a matrix of coefficients relating past values of the state vector to current values of the state vector. The stochastic trend representation in Moss and Shonkwiler consists of a measurement equation and two transition equations.

Past work by Bessler (1980), Goodwin and Ker (1998), Ker and Goodwin (2000) model yields using ARIMA processes. Goodwin and Ker (1998) evaluate a range of specifications using Akaike's information criterion and Schwarz's criterion. An ARIMA(0, 1, 2) specification is found to be optimal by both criteria. Ker and Goodwin (2000) provide a detailed discussion about the ARIMA(0, 1, q) model: It is a stochastic

linear trend with autocorrelated innovations. The linear trend represents technological advances while the moving average component suggests that the underlying factors generating the sequence of innovations have effects on future yields. Consider, for example, a drought in time t. It is obvious that yields in time t will be affected by drought conditions in time t. It is also reasonable that the soil in time t + 1, which influences yields in time t + 1, would have leftover effects from the drought in time t. Because weather is not a conditioning variate, these effects are represented in the innovations. Consequently, the innovations in time t would influence yields in time t + 1, thereby suggesting the existence of an MA component. Ker and Goodwin (2000) also point out a problem with estimating the ARIMA(0, 1, 1) model—the need to employ nonlinear least squares in small samples where convergence and parameter stability become issues. They replace the MA(1) process with an AR(4) representation to address the problem.

Ozaki et al., (2008) make use of Bayesian hierarchical models to capture the data generating process of Brazilian corn yields. They assume that the observed data follow a normal distribution, the mean of which is stochastic. To select among a large number of potential candidate models, a minimum mean square prediction error criterion is used. The stochastic mean component of the selected model includes both deterministic and stochastic trend. The deterministic part is a linear time trend and the stochastic part follows a first-order autoregressive model AR(1).

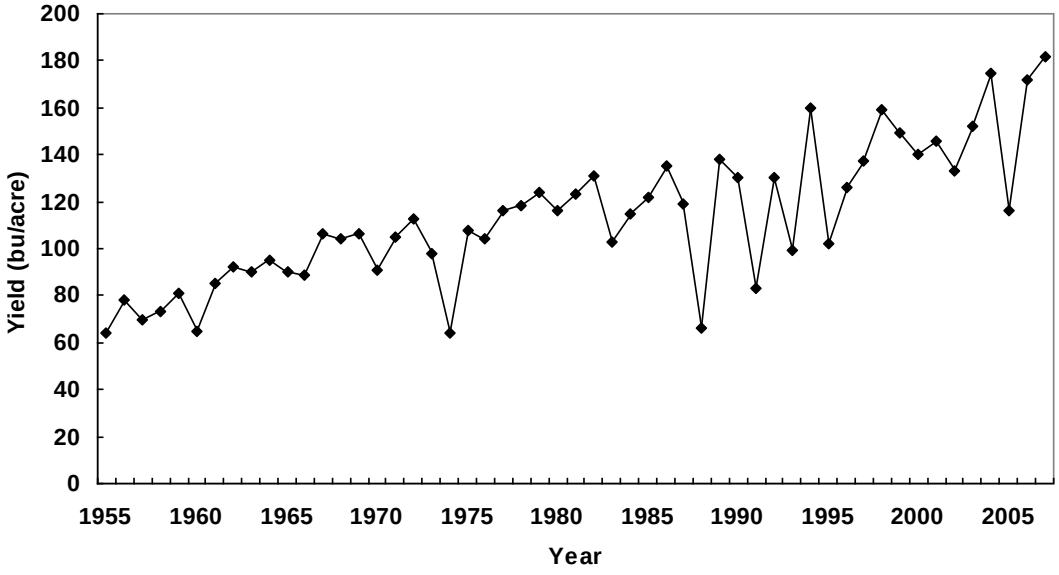
CHAPTER FOUR

DATA AND METHODOLOGY

4.1 Data and Heteroskedasticity Considerations

National Agricultural Statistics Service (NASS) county yields over the period 1955 to 2007 of 102 Illinois counties are used in the analysis. The crop analyzed is corn for grain. Figure 2 illustrates annual mean yields from three randomly sampled counties. All figures indicate yields have a slight upward trend with several years bearing extremely low yields.

It is ideal to predict mean yields by examining model selection criteria for each county-crop-practice combination and selecting the optimal specification and indexing parameters for each combination. Unfortunately, such an approach is operationally formidable given the large quantity of county-crop-practice combinations involved in GRP. The method outlined in this thesis considers less flexible but operationally more efficient way of predicting mean yields. Following the practice of FCIC, the central



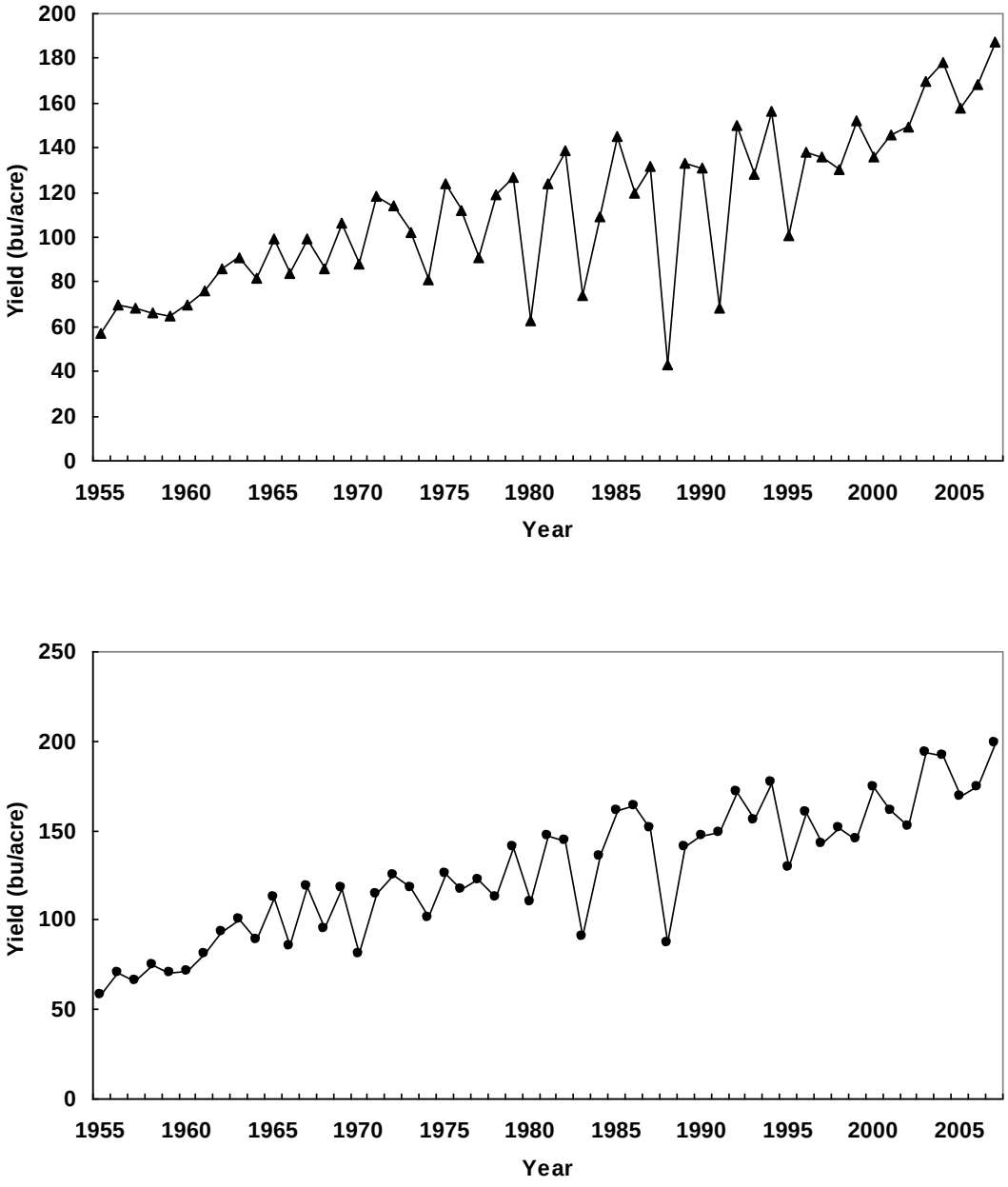


Figure 2: Plots of County Yield Data: Illinois Corn—All Practices

Source: USDA National Agricultural Statistics Service

tendency of yields is assumed to have a one-knot linear spline structure and the knot point is constrained to be equivalent across counties.

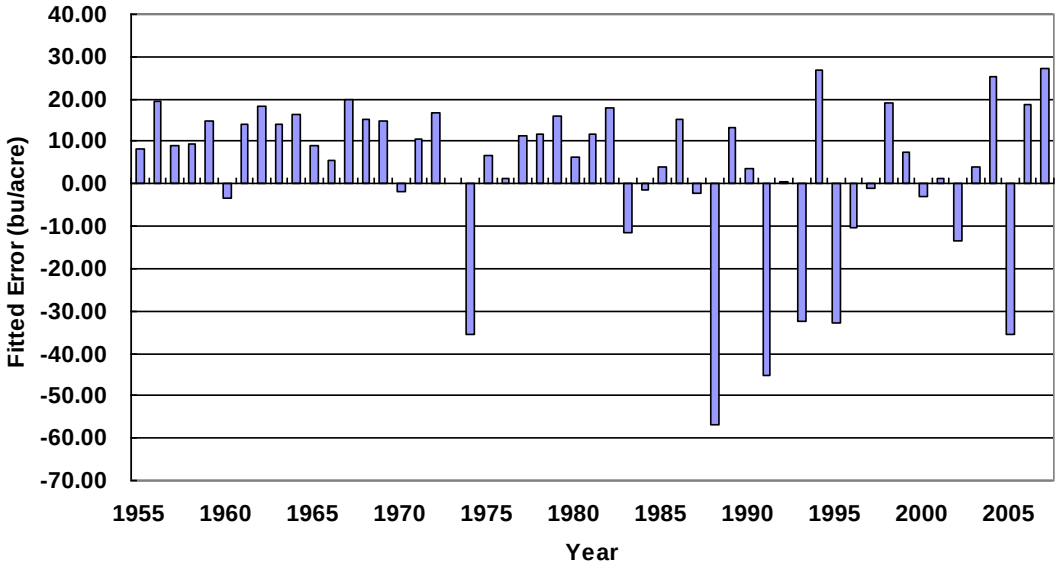
A concern with time series data is the existence of heteroskedasticity. The crop

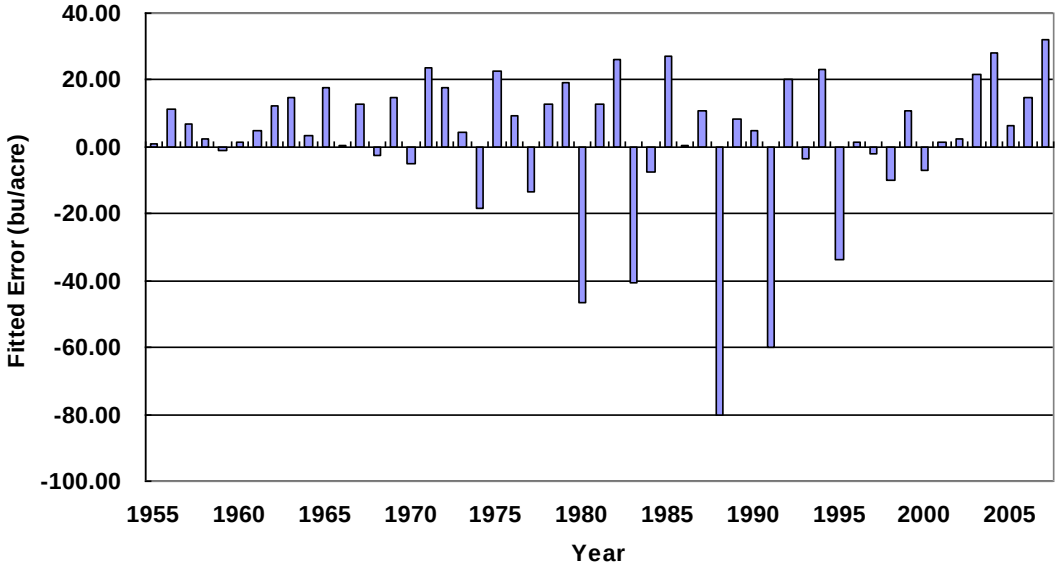
yield variance is possible to vary over time or with respect to some exogenous factors. For instance, yield variance may increase with yield levels suggesting yields have a constant coefficient of variation. When a constant coefficient of variation of crop yield is imposed, premium rates will be inflated if variance increases at a slower pace than that of the mean yield. Conversely, premium rates will be deflated if mean yield increases slower than the variance does. The assumption of the constant coefficient of variation can be evaluated by testing for heteroskedasticity in the fitted residuals of the spline model (detrended yields).

Breusch and Pagan (1979) devised a Lagrange multiplier test of the hypothesis that $\sigma_i^2 = \sigma^2 f(\alpha_0 + \boldsymbol{\alpha}'\mathbf{z}_i)$, where $\mathbf{z}_i$ is a vector of independent variables. The model is homoskedastic if $\boldsymbol{\alpha} = 0$. Under the null hypothesis of homoskedasticity, the test statistic LM has a limiting chi-squared distribution with degrees of freedom equal to the number of variances in $\mathbf{z}_i$. This test requires a set of independent variables to be specified as the base of the disturbance variance. Figure 3 shows the annual fitted residuals from the spline model of the same three counties sampled in Figure 2.

The variance of the fitted residuals (detrended yields) appears to vary across time and relate to yield levels. Since the one-knot linear spline model is a time trend model, yield levels and time are not independent variables. Thus, two Breusch-Pagan tests are conducted separately, using time and predicted yields as independent variable respectively. Then, another test is conducted on the fitted residuals standardized by the predicted value of yields. The first two hypotheses consider whether the fitted residuals are homoskedastic across time or with respect to the predicted yields. The third hypothesis considers whether the standardized residuals are homoskedastic versus

heteroskedastic. It has been argued that the Breusch-Pagan Lagrange multiplier test is sensitive to the assumption of normality. Koenker and Bassett (1982) suggest a modified statistic, which will have the same asymptotic distribution as the Breusch-Pagan statistic under normality. But absent normality, there is some evidence that it provides a more powerful test. The Koenker and Bassett statistic is also calculated in each of the three





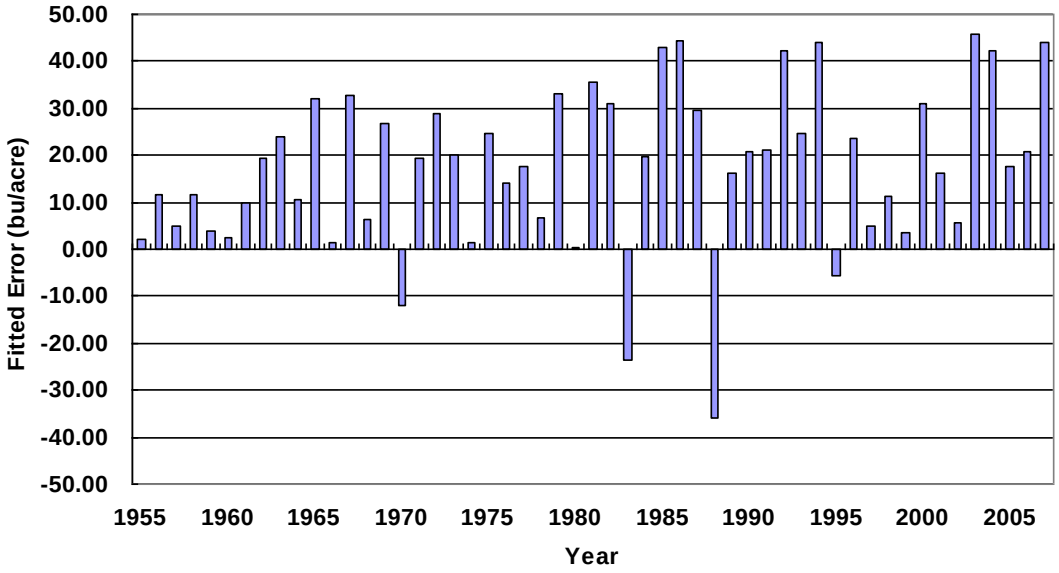


Figure 3: Plots of County Yield Fitted Residuals: Illinois Corn—All Practices

tests. A 5% level of significance is used for all tests. All test results are reported in Table 1. The tests strongly indicate that the raw residuals are heteroskedastic and the standardized residuals are homoskedastic, which implies that residuals have constant coefficient of variation.

Table 1: Heteroskedasticity Test Results: Counties Rejecting Null

		Hypothesis Test 1	Hypothesis Test 2	Hypothesis Test 3
Breusch-Pagan	Number	40	42	5
	Percent	39.22%	41.18%	4.90%
Koenker-Bassett	Number	37	37	7
	Percent	36.27%	36.27%	6.86%

Test 1 H_0: Raw residual is homoskedastic across time.

H_1: Raw residual is heteroskedastic.

Test 2 H_0: Raw residual is homoskedastic relative to predicted value.

H_1: Raw residual is heteroskedastic relative to predicted value.

Test 3 H_0: Residual standardized by predicted value is homoskedastic across time.

H_1: Residual standardized by predicted value is heteroskedastic across time.

4.2 Departures from Normality

As reviewed in Chapter 2, it is widely acknowledged that crop yields are not necessarily normal but may be negatively skewed. Normality tests are undertaken for each county. Take the concerns raised in Just and Weninger (1999) into account, the original yield series are made trend stationary by subtracting the fitted deterministic component. Since heteroskedasticity is detected, the fitted residuals are made variance stationary by dividing each data point by the predicted yield. Series with trend and heteroskedasticity removed are then tested on. The null assumes the standardized residuals are independent and identically distributed realizations from a Normal distribution. The alternative assumes the standardized residuals are realizations from a non-normal distribution.

The normal distribution is symmetric and mesokurtic. The symmetry implies that the third moment is zero. The standard measure of symmetry of a distribution is the skewness coefficient. Kurtosis is measure of the thickness of the tails of a distribution. The normal distribution is a usual yardstick for kurtosis; the mesokurtic value is the kurtosis of the normal distribution, which is 3. We can compare a distribution with the normal distribution by comparing its skewness with zero and its kurtosis to three (Greene, 2000). The Wald test statistic is:

$$(3)\ W = n\left[\frac{b_1^2}{6} + \frac{(b_2 - 3)^2}{24}\right] \xrightarrow{d} \chi_2^2$$

where b_1 is the measure of skewness and b_2 is the measure of kurtosis.

Normality is rejected in 63 counties or 61.76% of the counties with a 5% significance level, which suggest that yield distributions are not normal overall. In attempt to gain further information about the departure from normality, the skewness and kurtosis pairs of all 102 counties are plotted in Figure 4. The normal distribution corresponds to a point located at (0, 3) in the skewness-kurtosis plane. The sample points are not clustered around the normal point, which suggests that the yield distributions are not generated by a normal distribution. The corn samples display unanimous negative skewness and wide variation in the kurtosis. A negative skewness measure indicates negatively skewed distribution (mean < mode). Kurtosis measures less than three indicate thin tails relative to the normal distribution while kurtosis measures greater than three indicate fat tails relative to the normal distribution.

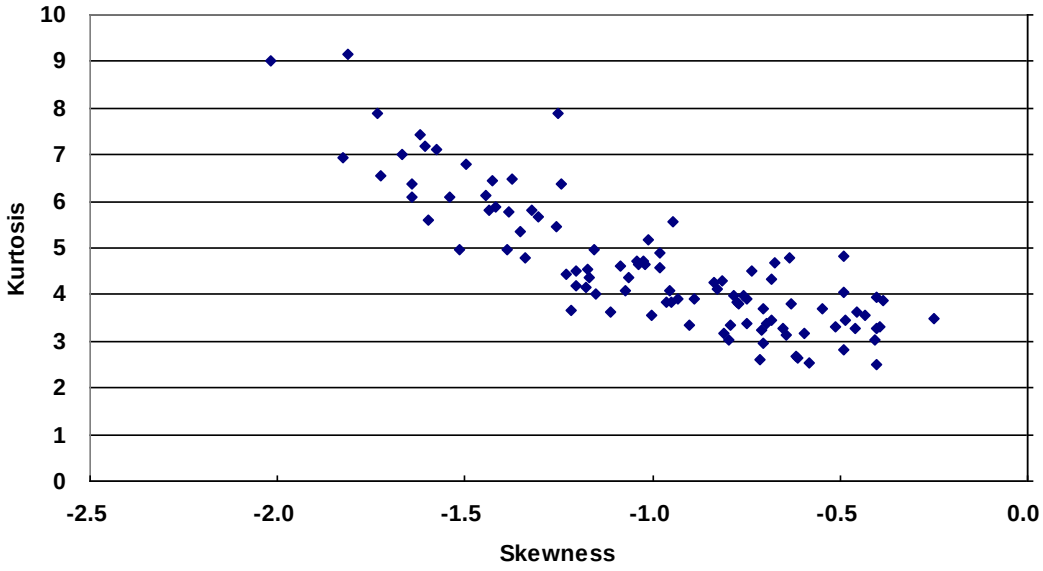


Figure 4: Skewness-Kurtosis Diagram for All Illinois Counties

Ker and Goodwin (2000) claims that the yield distributions may be approximated by two distinct sub-populations: a catastrophic sub-population and a non-catastrophic

sub-population. That is, in years when a catastrophic event occurs such as a drought, flood, freeze, etc., yields are drawn from the catastrophic sub-population. Conversely, in years when a catastrophic event does not occur, yields are drawn from the non-catastrophic sub-population. Thus, a mixture of two distributions can be used to approximate the residuals from the model, where the secondary distribution (from catastrophic years) lives on the lower tail of the primary distribution (from non-catastrophic years) and has significantly less mass. The secondary distribution would be expected to have less mass because catastrophic events are realized with far less frequency than their complement. Also, the secondary distribution would be expected to live on the lower tail of the primary distribution because realized yields tend to be far less in catastrophic years. Given this basic structure, mean yields may have a unimodal symmetric density (mass of catastrophic distribution is negligible), a negatively skewed density (mass of catastrophic distribution is non-negligible and distribution is relatively flat), or a negatively skewed bimodal density (mass of catastrophic distribution is non-negligible and distribution is relatively peaked).

In light of the prominent negative skewness found in the fitted residuals and the flexibility of the structure, a mixture of two normal distributions is assumed for the residuals from the model in this research.

4.3 Spline Model and Likelihood Functions

Given that we are using the GRP program as our background, we employ the temporal model used by FCIC in rating GRP contracts. The one-knot linear spline model can be specified as follows:

(4) $y_t = (\alpha_1 + \beta_1 t)\,\mathrm{I}_{(0,\psi)}(t) + (\alpha_2 + \beta_2 t)\,\mathrm{I}_{[\psi,\infty)}(t) + e_t$

s.t. $\alpha_1 + \beta_1\psi = \alpha_2 + \beta_2\psi$

where ψ is the knot point, t = 1, 2, …, 53 is the year, and I(.) is the indicator function. Therefore,

(5) $\alpha_1 + \beta_1\psi = \alpha_2 + \beta_2\psi \Rightarrow \alpha_2 = \alpha_1 + \beta_1\psi - \beta_2\psi$

Plug (5) into (4) and after arrangement, we obtain

(6) $y_t = \alpha_1 + \beta_1[t\,\mathrm{I}_{(0,\psi)}(t) + \psi\,\mathrm{I}_{[\psi,\infty)}(t)] + \beta_2(t-\psi)\,\mathrm{I}_{[\psi,\infty)}(t) + e_t$

To limit the dimension of the estimator, the parameters indexing the spline model are assumed not to vary among counties. The advantages of this assumption are a significant gain in degrees of freedom and more efficient estimation under the null. Define $\beta_0 \equiv \alpha_1$, $x_{1t} \equiv t\,\mathrm{I}_{(0,\psi)}(t) + \psi\,\mathrm{I}_{[\psi,\infty)}(t)$ and $x_{2t} \equiv (t-\psi)\,\mathrm{I}_{[\psi,\infty)}(t)$, then

(7) $y_{it} = f(\mathbf{x_t}, \boldsymbol{\beta}) + e_{it} = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + e_{it}$

where f is the one-knot linear spline temporal process, $\boldsymbol{\beta}$ is a 3×1 parameter vector, and e_{it} denotes the intra-county disturbance associated with the t^{th} realization in county i. Note that f, $\boldsymbol{\beta}$ and $\mathbf{x_t}$ are not indexed by county and are constant for all $i \in \mathbf{I}$, $\mathbf{I}$ = {1, 2, …, 102}. The total number of observations is $\mathbf{N} = \mathbf{I} \times \mathbf{T} = 102 \times 53 = 5406$. The heteroskedasticity results of section 3.1 indicate that disturbances have constant coefficient of variation. Therefore,

(8) $e_{it} = \varepsilon_{it} f(\mathbf{x_t}, \boldsymbol{\beta})$

where ε_{it} has constant variance, so that $e_{it} / f(\mathbf{x_t}, \boldsymbol{\beta})$ is homoskedastic. Specification of the distribution of the disturbance sequence will complete the characterization of the likelihood of county i.

The distribution describing the disturbance sequence, $\{e_{it}\}$, must be sufficiently flexible to accommodate conditional distributions for all counties. A mixture of two normal distributions is assumed for the disturbance sequence. Because of the necessity to limit the dimension of the estimator, certain restrictions are imposed. The parameters that vary by county are the variance parameters of the two normal distributions (σ_i^2, δ_i^2). The mean parameters of the two normal distributions (μ, η) and the mixing parameter α will be fixed across counties.

So for a given county *i*, the probability density function of e_{it} is

$$(9)\quad \mathrm{p}_e(e_{it}) = \alpha N(\mu, \sigma_i^2 f(\mathbf{x_t}, \boldsymbol{\beta})^2) + (1-\alpha) N(\eta, \delta_i^2 f(\mathbf{x_t}, \boldsymbol{\beta})^2)$$

where $N(\mu, \sigma^2)$ denotes the density of a normal distribution with mean μ and variance σ^2. The joint density of $\{e_{it}\}$ is

$$(10)\quad \mathrm{p}_e(e_{i1}, ..., e_{iT} \mid \mathbf{x_t}, \boldsymbol{\beta}, \mu, \eta, \alpha, \sigma_i, \delta_i) = \prod_{t=1}^{T} \mathrm{p}_e(e_{it})$$

Thus, the joint density of yields for county *i* is

$$\begin{aligned} &\mathrm{p}(y_{i1}, ..., y_{iT} \mid \mathbf{x_t}, \boldsymbol{\beta}, \mu, \eta, \alpha, \sigma_i, \delta_i) \\ (11)\quad &= \prod_{t=1}^{T} \mathrm{p}(y_{it}) \\ &= \prod_{t=1}^{T} [\alpha N(\mu + f(\mathbf{x_t}, \boldsymbol{\beta}), \sigma_i^2 f(\mathbf{x_t}, \boldsymbol{\beta})^2) + (1-\alpha) N(\eta + f(\mathbf{x_t}, \boldsymbol{\beta}), \delta_i^2 f(\mathbf{x_t}, \boldsymbol{\beta})^2)] \end{aligned}$$

This density can be obtained by substituting $e_{it} = y_{it} - f(\mathbf{x_t}, \boldsymbol{\beta})$ into $\mathrm{p}_e(e_{i1}, ..., e_{iT})$ because the Jacobian of $(e_{i1}, ..., e_{iT})$ with respect to $(y_{i1}, ..., y_{iT})$ is the identity matrix of order *T*.

Parameter vector $\mathbf{v_i} = \{\sigma_i, \delta_i\}$ of the disturbance distribution varies from county to county. Part of this inter-county variation can be explained by systematic dependence on county specific attributes. The unexplained portion of the inter-county variation is

assumed to be random and is characterized by a probability density h. The interaction of systematic and random sources of variation is represented by an inter-county regression model for each parameter in $\boldsymbol{\nu}_{\mathbf{i}}$.

$$(12)\quad \boldsymbol{\nu}_{\mathbf{ki}} = g_k(\boldsymbol{\omega}_{\mathbf{i}}, \boldsymbol{\gamma}, \mathbf{z}_{\mathbf{ki}}) \qquad k = 1, 2$$

where $\boldsymbol{\omega}_{\mathbf{i}}$ is a vector of county specific attributes, $\boldsymbol{\gamma}$ is a vector of unknown fixed effects, and $\mathbf{z}_{\mathbf{i}} = (z_{1i}, ..., z_{Mi})$ is an M-dimensional vector of inter-county random effects with density h(z). Determination of an appropriate model for g(.) involves choice of county specific attributes to make up $\boldsymbol{\omega}_{\mathbf{i}}$ and specification of a mathematical form to represent the nature of the dependence of $\boldsymbol{\nu}_{\mathbf{i}}$ on $\boldsymbol{\omega}_{\mathbf{i}}$ and $\mathbf{z}_{\mathbf{i}}$. Collecting all the fixed parameters into a vector $\boldsymbol{\tau} = (\boldsymbol{\gamma}, \boldsymbol{\beta}, \alpha, \mu, \eta)$, the population likelihood may be written as

$$(13)\quad L(\boldsymbol{\tau}, h) = \prod_{i=1}^{N} \int p(y_{i1}, ..., y_{iT} \mid \mathbf{x}_{\mathbf{t}}, \boldsymbol{\omega}_{\mathbf{i}}, \boldsymbol{\tau}, \mathbf{z}_{\mathbf{i}})\, h(z)\, dz$$

Here, the form of p(.) differs from that given in equation (11) because $\boldsymbol{\nu}_{\mathbf{ki}} = g_k(\boldsymbol{\omega}_{\mathbf{i}}, \boldsymbol{\gamma}, \mathbf{z}_{\mathbf{ki}})$ has been substituted to emphasize the dependence of the density on $\boldsymbol{\omega}_{\mathbf{i}}$ and $\mathbf{z}_{\mathbf{i}}$. The log likelihood is

$$(14)\quad l(\boldsymbol{\tau}, h) = \sum_{i=1}^{N} \log \int p(y_{i1}, ..., y_{iT} \mid \mathbf{x}_{\mathbf{t}}, \boldsymbol{\omega}_{\mathbf{i}}, \boldsymbol{\tau}, \mathbf{z}_{\mathbf{i}})\, h(z)\, dz$$

The primary objective is estimation of and inference regarding the random effects density h($\mathbf{z}$) and the fixed parameters $\boldsymbol{\tau} = (\boldsymbol{\gamma}, \boldsymbol{\beta}, \alpha, \mu, \eta)$. Once $\boldsymbol{\tau}$ and h(z) are determined, the individual parameters $\boldsymbol{\nu}_{\mathbf{i}} = \{\sigma_i, \delta_i\}$ can be estimated by empirical Bayes. We estimate h(z) nonparametrically, simultaneously with $\boldsymbol{\tau}$, by maximizing the population likelihood. The procedure is described in the next section.

4.4 SNP Estimators

Nonparametric estimation of h(z) allows one to detect unusual features of the population such as multimodality or excess dispersion, which often indicates the presence of systematic inter-county variability and the need for a more refined inter-county regression model g(.). It also affords protection against incorrect assumption regarding h(z) that can bias estimates of $\boldsymbol{\tau}$ and lead to erroneous inferences (Davidian and Gallant, 1992).

Seminonparametric (SNP) maximum likelihood method, by sacrificing some generality in favor of a smoothness assumption, estimates the density of the random effects nonparametrically, jointly with the fixed effects, and inference is possible. The method uses a series expansion that follows from smoothness assumptions to represent the density, due to Gallant and Nychka (1987), and uses quadrature to compute the likelihood. Standard algorithms are used for optimization. Empirical Bayes estimates of random parameters are obtained by computing posterior modes. This approach is taken from the pharmacokinetics literature where it has been advanced by Davidian and Gallant in modeling the pharmacokinetics of quinidine and the pharmacokinetics of phenobarbital (1992, 1993). The contents of this section draw heavily from these articles. Fortran code for the nonlinear mixed effects model with SNP estimation of the fixed parameters and random effects density is written by Davidian and Gallant[7]. Additional Fortran code that is written for this article is located in Appendix.

SNP estimation of $\boldsymbol{\tau}$ and h(z) maximize the likelihood function (14). Maximizing likelihood function (14) is equivalent to minimizing over $\boldsymbol{\tau}$ and h

[7] The Fortran code is available via "http://econ.duke.edu/webfiles/arg/nlmix/".

$$(15)\ \mathrm{s_N}(\boldsymbol{\tau}, \mathrm{h}) = -\frac{1}{N}\sum_{i=1}^{N}\log\int \mathrm{p}(y_{i1},...,y_{iT} \mid \mathbf{x_t}, \boldsymbol{\omega_i}, \boldsymbol{\tau}, \mathbf{z_i})\,\mathrm{h}(z)\,\mathrm{d}\,z$$

After the estimates $(\hat{\boldsymbol{\tau}}, \hat{\mathrm{h}})$ are obtained, empirical Bayes estimates of the random effects $\mathbf{z_i}$ are computed as the values $\hat{\mathbf{z}}_\mathbf{i}$ that maximize with respect to $\mathbf{z}$

$$(16)\ \mathrm{p}(y_{i1},...,y_{iT} \mid \mathbf{x_t}, \boldsymbol{\omega_i}, \hat{\boldsymbol{\tau}}, \mathbf{z_i})\,\hat{\mathrm{h}}(z)$$

Then, the empirical Bayes estimate of a county's disturbance parameter $\boldsymbol{\nu}_\mathbf{i}$ can be obtained by plugging in the estimated parameters into g(.), which is $\hat{\boldsymbol{\nu}}_\mathbf{i} = \mathrm{g}(\boldsymbol{\omega_i}, \hat{\boldsymbol{\gamma}}, \hat{\mathbf{z}}_\mathbf{i})$.

The approach is based on the assumption that the true density h(.) belongs to a class of smooth densities $\mathcal{H}$ described below. A mathematical description of the class $\mathcal{H}$ is given in Gallant and Nychka (1987). The main requirement is that a density h(.) in $\mathcal{H}$ must satisfy the smoothness restriction that h(.) be at least M/2 times differentiable. As a consequence, densities exhibiting unusual behavior such as kinks, jumps, or oscillation are excluded from consideration; however densities in $\mathcal{H}$ may be skewed, multimodal, and fat-tailed or thin-tailed relative to the M-variate normal density. Thus, the assumption that true density belongs to $\mathcal{H}$ rules out densities with unusual features that are unlikely to characterize the random effects while allows for a wide range of behaviors.

A density from $\mathcal{H}$ can be represented as an infinite series expansion. The format of the expansion is

$$(17)\ \mathrm{h}(z) = \frac{[\mathrm{P}(z)]^2\,\phi_M(z)}{\int [\mathrm{P}(z)]^2\,\phi_M(z)dz}$$

where P(z) is a Hermite expansion and $\phi_M(z)$ denotes the multivariate normal density of dimension M. P(z) is squared to insure non-negativity. The denominator is to insure that the estimated density h(z) integrates to one. Given the division, h(z) is homogeneous

function of the coefficients of the expansion P(z) and hence may only be determined to within a scalar multiple. Unique representation may be achieved by setting the constant term of the expansion to one.

For practical application, one may consider approximation of h(z) by a truncation of this expansion to finite number of leading terms. By increasing the truncation point when specification error is detected, these procedures have nonparametric properties in that they are consistent estimators. The truncated expansion is

$$(18)\quad \mathrm{h}_{\lambda}(z) = \frac{[\mathrm{P}_{\lambda}(R^{-1}z)]^2 N_M(z \mid 0, RR')}{\int [\mathrm{P}_{\lambda}(u)]^2 N_M(u \mid 0, I)du}$$

where $N_M(\cdot \mid \boldsymbol{\mu}, \sum)$ denotes the multivariate normal density of dimension M with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\sum$ and R is an upper-triangular matrix. $\mathrm{P}_{\lambda}(z)$ is a polynomial that is the sum of all powers and cross-products of the components of z up to degree λ. For illustration, if M = 2 and λ =2 then z = (z_1, z_2), and

$$(19)\quad P_2(z) = a_{00} + a_{10}z_1 + a_{01}z_2 + a_{20}z_1^2 + a_{02}z_2^2 + a_{11}z_1z_2$$

The coefficients of the polynomial are $(a_{00}, a_{10}, a_{01}, a_{20}, a_{02}, a_{11})$ with $a_{00} \equiv 1$. Let $\theta_{(1)}$ be a vector whose elements are the coefficients of P_{λ} and let $\theta_{(2)}$ be a vector of dimension M(M+1)/2 whose elements are the elements of the upper-triangular matrix R stacked column by column. Let $\theta = (\theta_{(1)}, \theta_{(2)})$, vector θ completely describes the truncated expansion and the dimension of θ is determined by the degree λ of the polynomial and dimension M of the random effects z. In the example, with M = 2 and λ =2, the dimension of θ is 6 + 3 = 9 and the denominator of (18) is a weighted sum of products of moments to the fourth order of the standard normal distribution.

Once h(z) has been approximated by the truncated expansion $h_\lambda(\cdot|\theta)$, estimation of $\boldsymbol{\tau}$ and h(z) becomes a standard finite dimensional nonlinear optimization problem. One minimizes $s_N[\boldsymbol{\tau}, h_\lambda(\cdot|\theta)]$ in the variables $\boldsymbol{\tau}$ and θ to obtain $\hat{\boldsymbol{\tau}}$ and $\hat{\theta}$. The estimate of h is then $h_\lambda(\cdot|\hat{\theta})$. As long as the degree λ of the polynomial increases with the sample size N, the estimates $(\hat{\boldsymbol{\tau}}, \hat{\theta})$ obtained in this way are consistent estimators. The appropriate truncation point λ is chosen by visual inspection of the estimated density of the random effects.

CHAPTER FIVE

FINDINGS AND SUMMARY

5.1 Empirical Results

Results from least squares estimates are used as initial starting values. Because there are many local maximum, numerous starting values and random number seeds are tried. To select the proper knot point, a grid search is conducted, namely twenty models are estimated with the knot point ranging from 1960 to 1980 respectively and the likelihood values of those twenty models are compared. The maximum likelihood occurs with a knot point of 1969, thus it is selected as the estimated knot point.

Two caveats need to be specified regarding the estimation of the population likelihood. First, not all parameters are estimable. In the common situation of a linear model with normal disturbances, it is obvious that the location parameter of the disturbance distribution, denoted μ, and the intercept term from the model, denoted β, are non-estimable. However, $\mu + \beta$ is estimable. In general, the location parameter of the disturbance distribution is restricted to zero ($\mu = 0$) and hence the intercept parameter is recovered as $\beta = \mu + \beta$. Now consider the situation where the disturbance distribution is a mixture of two normal distributions with non-zero location parameters (μ, η) and non-zero intercept β_0. Without any restrictions μ, η and β_0 are non-estimable while $\mu + \beta_0$ and $\eta + \beta_0$ are estimable. Imposing the restriction that the disturbance distribution has zero mean, that is $\alpha\mu + (1-\alpha)\eta = 0$, then μ, η and β_0 are recoverable from the estimates of $\mu + \beta_0$ and $\eta + \beta_0$. Specifically,

$$
\begin{aligned}
&\alpha(\mu+\beta_0)+(1-\alpha)(\eta+\beta_0)=c \\
&\alpha\mu+\alpha\beta_0+(1-\alpha)\eta+(1-\alpha)\beta_0=c \\
&\alpha\mu+(1-\alpha)\eta+\beta_0=c \\
&\alpha\mu+(1-\alpha)\eta=0\Rightarrow\beta_0=c
\end{aligned} \tag{20}
$$

As a result, $\mu=\mu+\beta_0-c$ and $\eta=\eta+\beta_0-c$. This approach is used to recover $\hat{\mu}$, $\hat{\eta}$ and $\hat{\beta}_0$.

Second, unlike in Davidian and Gallant (1992, 1993), the model selection criteria (AIC, BIC, HQ) are relatively uninformative. In Davidian and Gallant applications, the parameters of the non-random component of the model vary across experimental units while the parameters indexing the disturbance distribution remain constant. In this application, the parameters of the non-random component of the model are constant whereas the parameters indexing the disturbance distribution vary across counties. As a consequence, changes in g(.) and h(.) will not significantly influence the likelihood and hence not significantly influence the model selection criteria. As a result, an alternative methodology must be used to recover g(.) and h(.).

Methodology outlined in Davidian and Gallant (1992, 1993) is used to identify an appropriate model given the available county specific attributes. At first, the models are fit without any county specific attributes and the truncation point λ is increased until the empirical Bayes estimates of the random effects $\hat{\mathbf{z}}_i$ separate. Davidian and Gallant (1992) note that the occurrence of this separation indicates the truncation degree λ is large enough. Also noted by Davidian and Gallant (1992) is that omission of influential county specific attributes from the inter-county regression function g(.) will tend to yield a fat-tailed or multimodal estimate of the random effects density h(z). These facts are exploited to identify an appropriate truncation point and relevant county specific attributes.

The following log-linear inter-county regression functions without any county specific attributes are initially estimated:

$$(21)\quad \begin{aligned} \sigma_i &= \exp(\gamma_1 + z_{1i}) \\ \delta_i &= \exp(\gamma_3 + z_{2i}) \end{aligned}$$

The log-linear form ensures positivity of standard deviation parameters. Figure 5 plots the estimated joint density of the random effects for $\lambda = 2$. Note the separation suggests an expansion of order $\lambda = 2$ is sufficient. National Agricultural Statistics Service (NASS) keeps data of harvested acreage of every county each year. So both the mean acreage and standard deviation of the acreage appear to be the natural candidates for county specific attributes. The empirical Bayes estimates of the inter-county random effects are plotted against available county attributes in Figure 6. However, Figure 7 suggests that these two attributes are highly correlated[8]. Thus, only mean acreage is used as the county specific attribute.

[8] The correlation coefficient between mean acreage and standard deviation of acreage is 0.915.

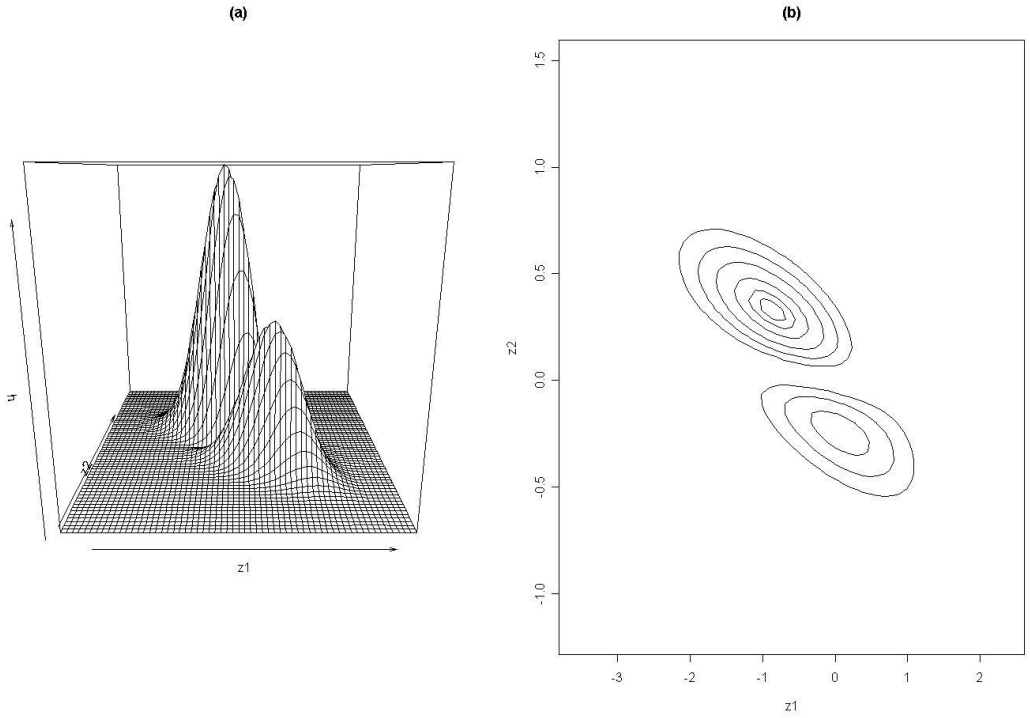
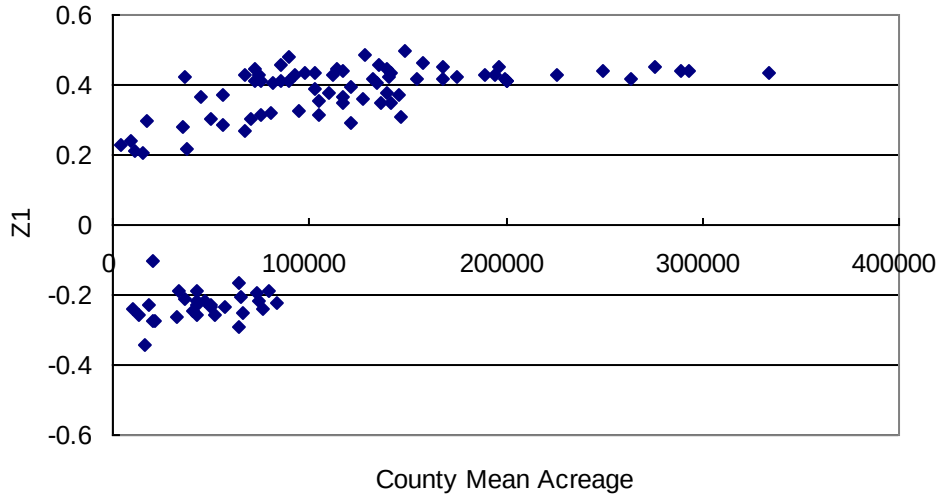
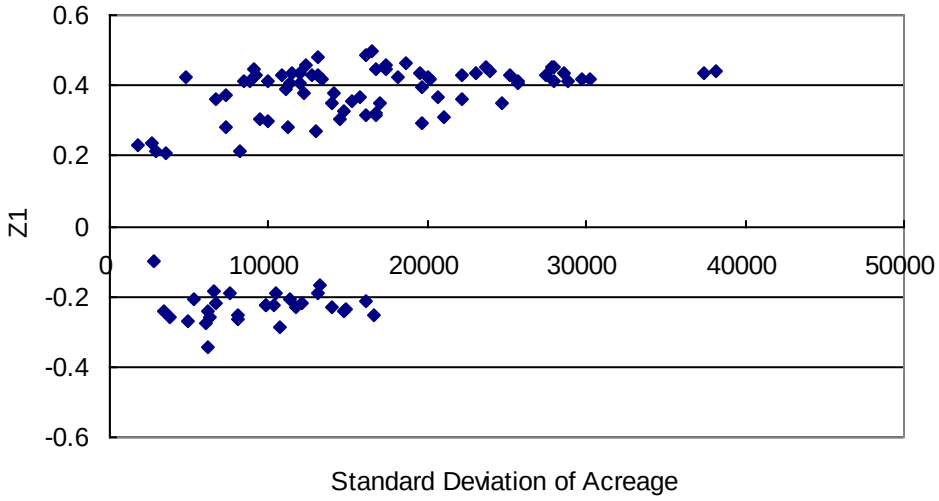
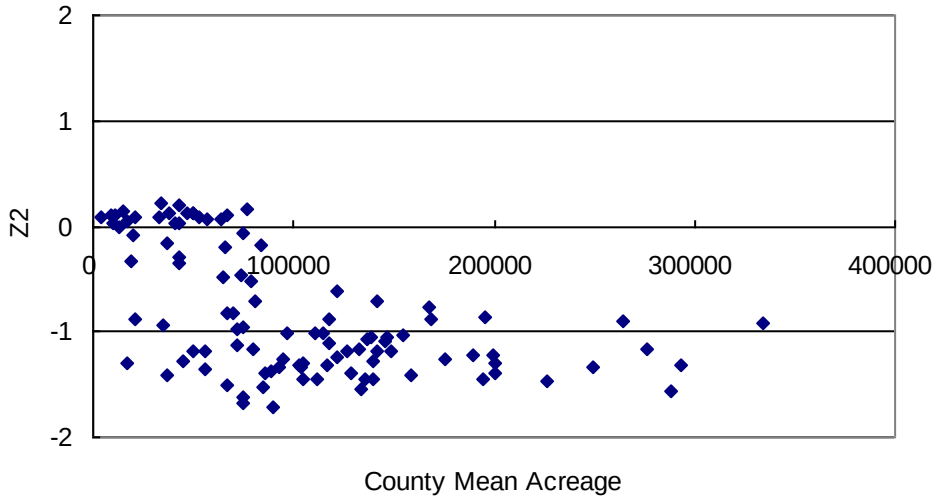


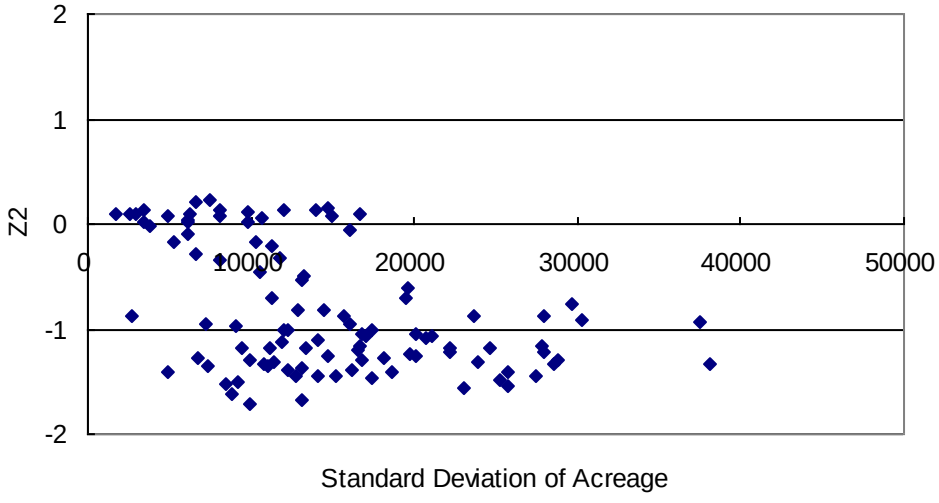
Figure 5: Estimated Inter-County Random Effects Density: No County Specific Attributes. (a) Perspective Plot of the Estimated Joint Density; (b) Contour Plot of the Estimated Joint Density at Quantiles 10%, 25%, 50%, 75%, 90%, 95%.

Figure 6: Inter-County Regression Graphics: No County Specific Attributes




0.6
0.4
0.2
0
-0.2
-0.4
-0.6
Z1
0
10000
20000
30000
40000
50000
Standard Deviation of Acreage


2
1
0
-1
-2
Z2
0
100000
200000
300000
400000
County Mean Acreage


2
1
0
-1
-2
Z2
0
10000
20000
30000
40000
50000
Standard Deviation of Acreage

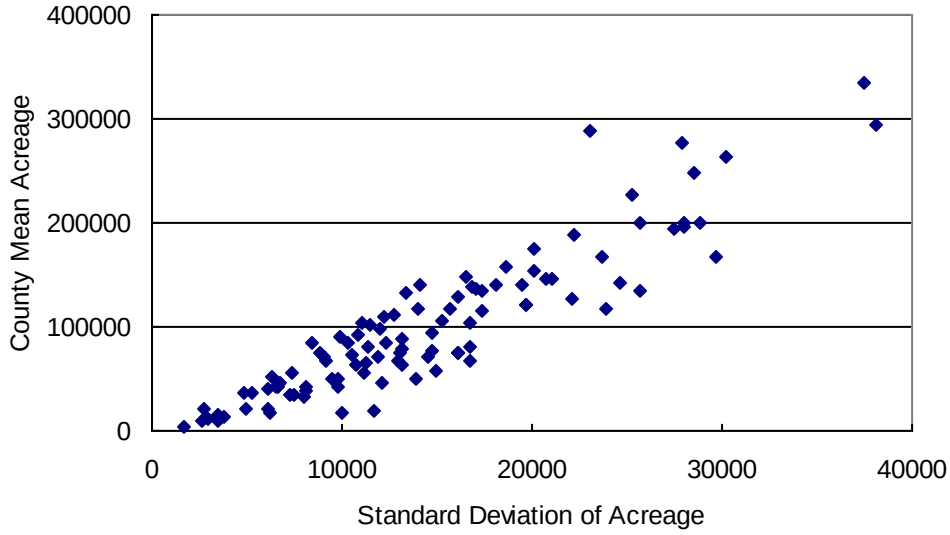


Figure 7: Mean Acreage versus Standard Deviation of Acreage

The following inter-county regression models are estimated:

$$(22)\quad \begin{aligned} \sigma_i &= \exp(\gamma_1 + \gamma_2(\text{mean acreage})_i + z_{1i}) \\ \delta_i &= \exp(\gamma_3 + \gamma_4(\text{mean acreage})_i + z_{2i}) \end{aligned}$$

Using this model, Figure 5 and Figure 6 are reproduced in Figure 8 and Figure 9 respectively. Ideally, after the county specific attributes are included in the model, the separation of the joint density of the random effects will disappear. However, the separation still exists in Figure 8. One possible explanation is that some influential county specific attribute is omitted. Several possible attributes (e.g. weather factors) are tried but all fail to eliminate the separation. Should this separation be due to the omission of some influential factor, it would decrease the efficiency of the model to some extent. However, since the deficiency occurs at the regression models for standard deviations, it would only affect the second moment of the yield density moderately.

The separation can also be a result of the inflexible model specification. The central tendency is constrained to be the same for all counties. If the estimated central

tendency systematically overestimates or underestimates the yields of a group of counties, the systematic fitted errors will be reflected in the estimates of the disturbance terms. In addition, only the variance parameters can differ by county, thus the variance parameter estimates of this group of counties must be fundamentally different from those of the other counties. And the groupwise difference cannot be explained by any county specific attributes. As a result, the estimated joint density of the random effects becomes separated to accommodate the groupwise difference. In this case, the separation is an artifact due to the inflexible model specification and cannot be eliminated by adding county specific attributes in the inter-county regression models.

Also note in Figure 8 that the joint density is thin suggesting high correlation between the two standard deviation parameters. This is not surprising as the parameters both measure dispersion. Also of interest is that the random effects plotted against the attributes in Figure 9 appear to have flatter slopes and are closer to zero than in Figure 6.

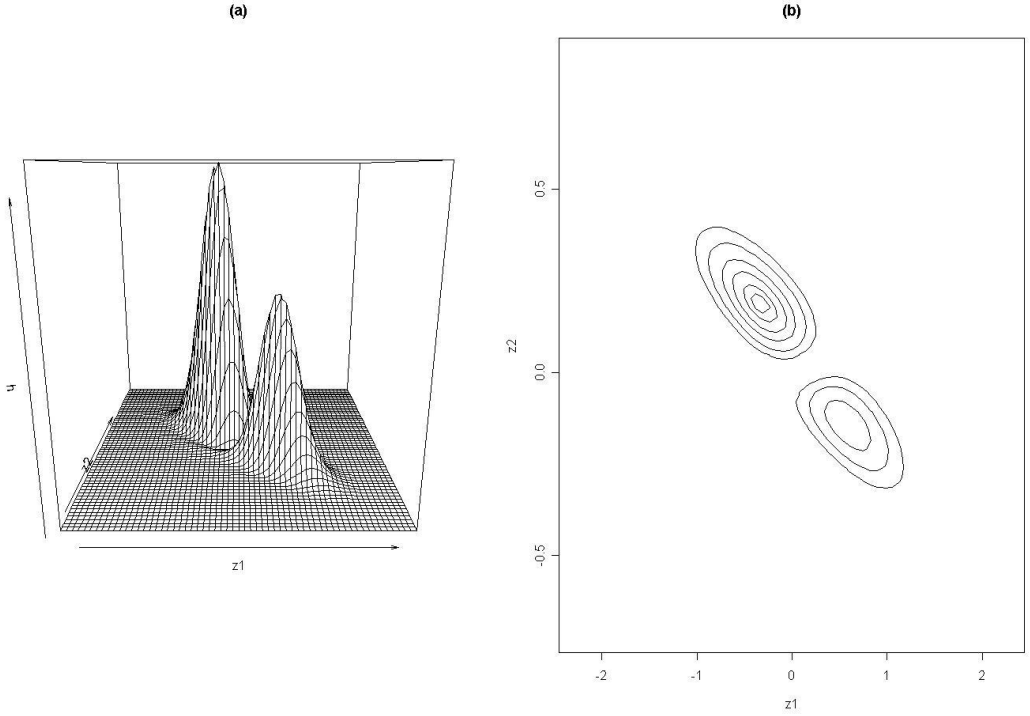
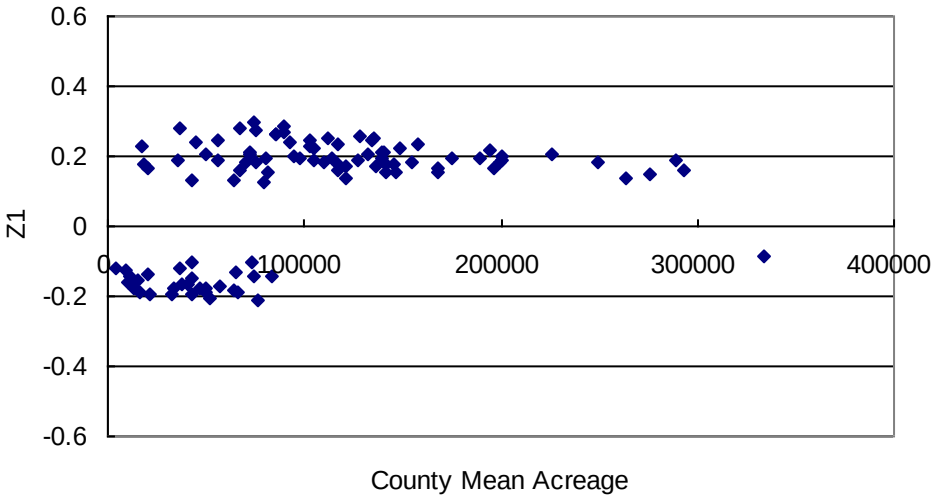
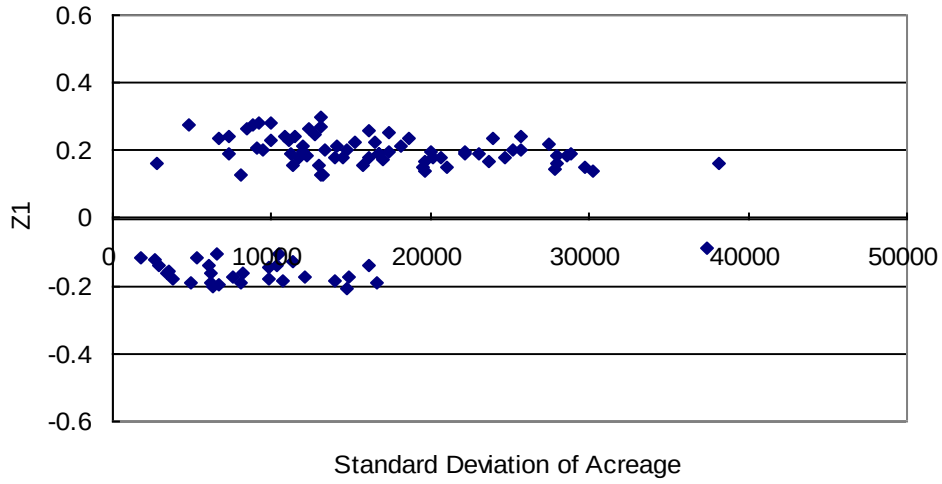
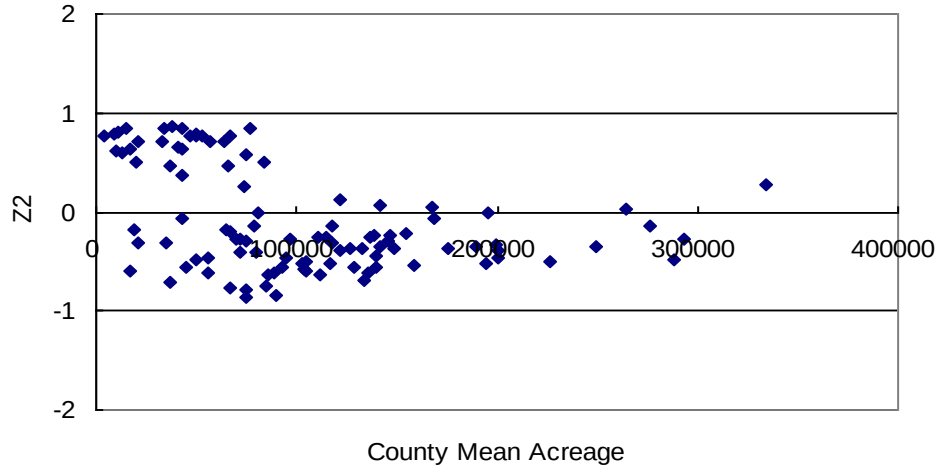


Figure 8: Estimated Inter-County Random Effects Density: With County Specific Attributes. (a) Perspective Plot of the Estimated Joint Density; (b) Contour Plot of the Estimated Joint Density at Quantiles 10%, 25%, 50%, 75%, 90%, 95%.

Figure 9: Inter-County Regression Graphics: With County Specific Attributes




0.6
0.4
0.2
0
-0.2
-0.4
-0.6
Z1
0
10000
20000
30000
40000
50000
Standard Deviation of Acreage


2
1
0
-1
-2
Z2
0
100000
200000
300000
400000
County Mean Acreage

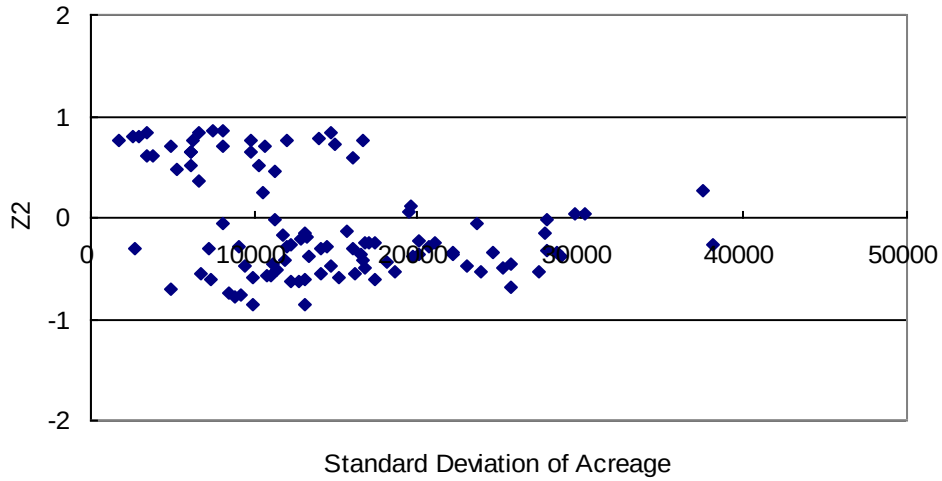

2
1
0
-1
-2
Z2
0
10000
20000
30000
40000
50000
Standard Deviation of Acreage

Table 2 contains the parameter estimates and standard deviations from the estimation results. All estimates are significant at 5% level. The fact that $\hat{\beta}_2$ is smaller than $\hat{\beta}_1$ indicates that the yield increases at a smaller rate after 1969. Intuitively, variance parameters decrease as the acreage increase in accordance with the law of large numbers, thus both γ_2 and γ_4 ought to be negative. However, $\hat{\gamma}_4$ is positive which may be caused by omission of the influential explanatory variable. The second normal distribution is located on the lower tail of the first normal distribution maintaining roughly 35% of the mass, suggesting that these yield distributions tend to be more skewed as opposed to bimodal. What is unexpected is that the variances of the second normal distribution are not uniformly higher than those of the first normal distribution, implying not all 102 counties have negatively skewed yield distribution. 29 out of 102 (28.4%) counties yield a positively skewed distribution. Figure 10 shows estimated standard deviations $(\hat{\sigma}_i, \hat{\delta}_i)$ of 102 counties. Two clusters of points are observed: the upper left cluster corresponds to counties where the variance of the second normal distribution is higher than the first normal distribution; the lower right cluster represents the 29 counties which yield unexpected positively skewed distributions.

Table 2: Parameter Estimates from SNP Maximum Likelihood Estimation

Parameter	Estimate	Standard Deviation
β_0 *	53.4219	
β_1	2.5157	0.0560
β_2	1.6758	0.0305
γ_1	-1.6646	0.1213
γ_2	-0.0163	0.0090
γ_3	-1.9220	0.0590
γ_4	0.0239	0.0052
$\mu + \beta_0$	62.4719	0.4624
μ *	9.0500	0.4624
$\eta + \beta_0$	36.6191	0.5326
η *	-16.8028	0.5326
α	0.6499	0.0088
Var(z_1)	0.3206	0.1067
Var(z_2)	0.0376	0.0187
$\rho(z_1, z_2)$	-0.9134	

* Parameter estimate recovered given zero mean restriction on disturbance distribution.

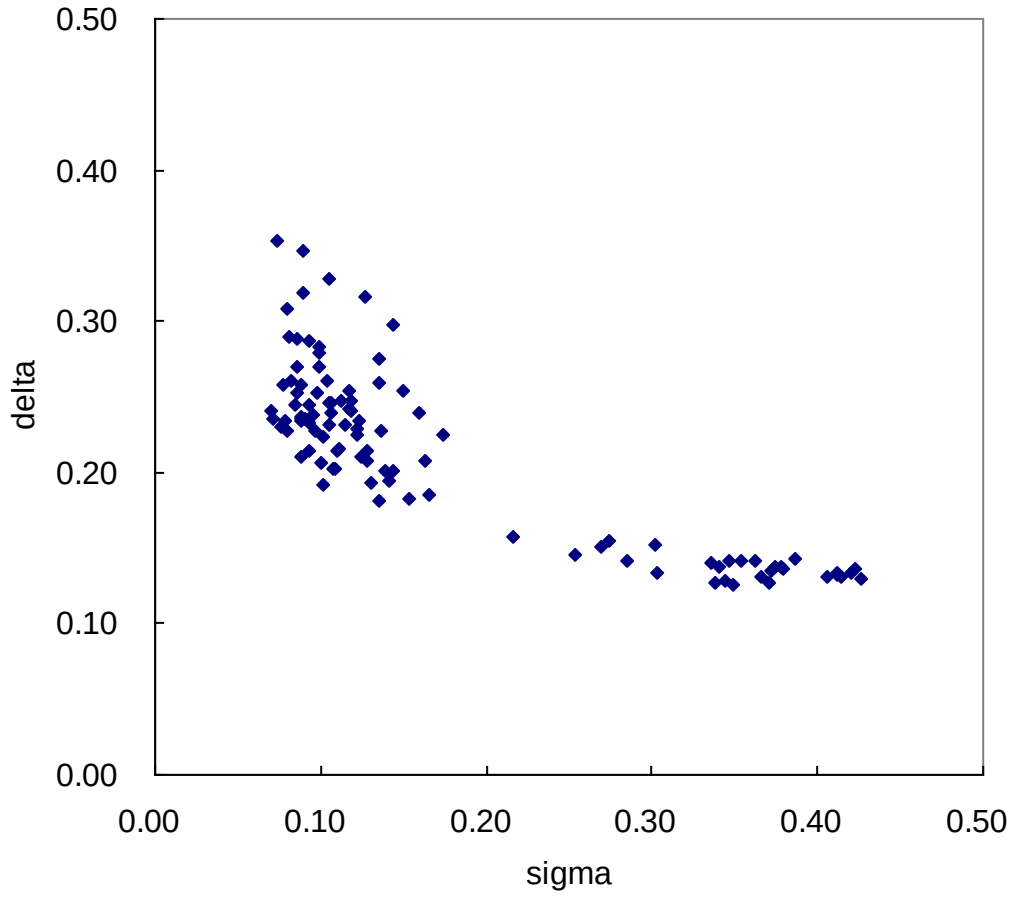
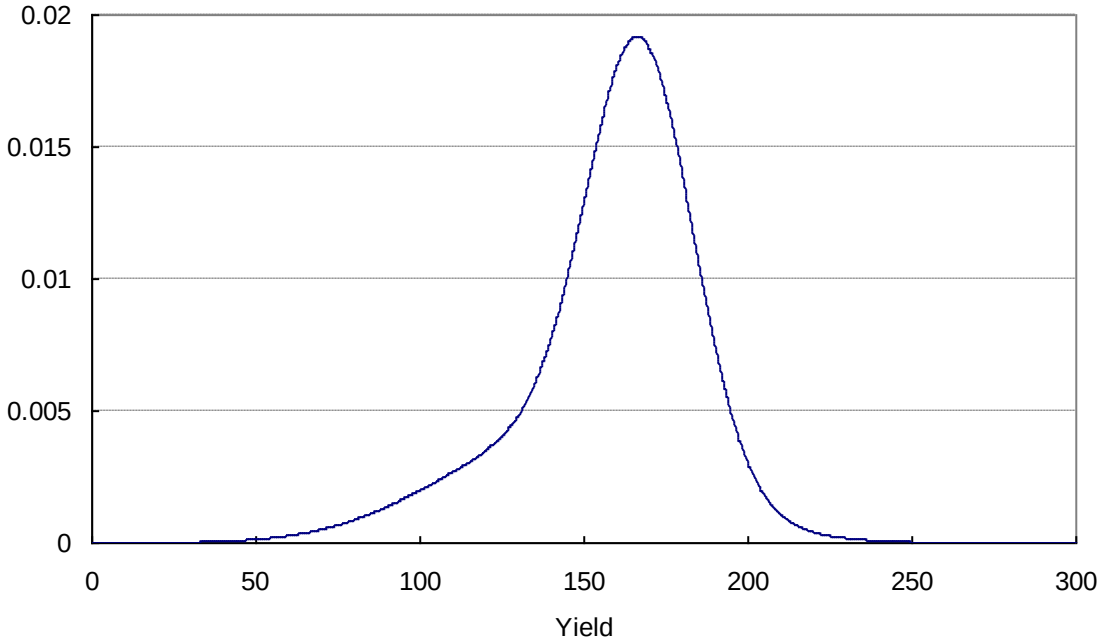
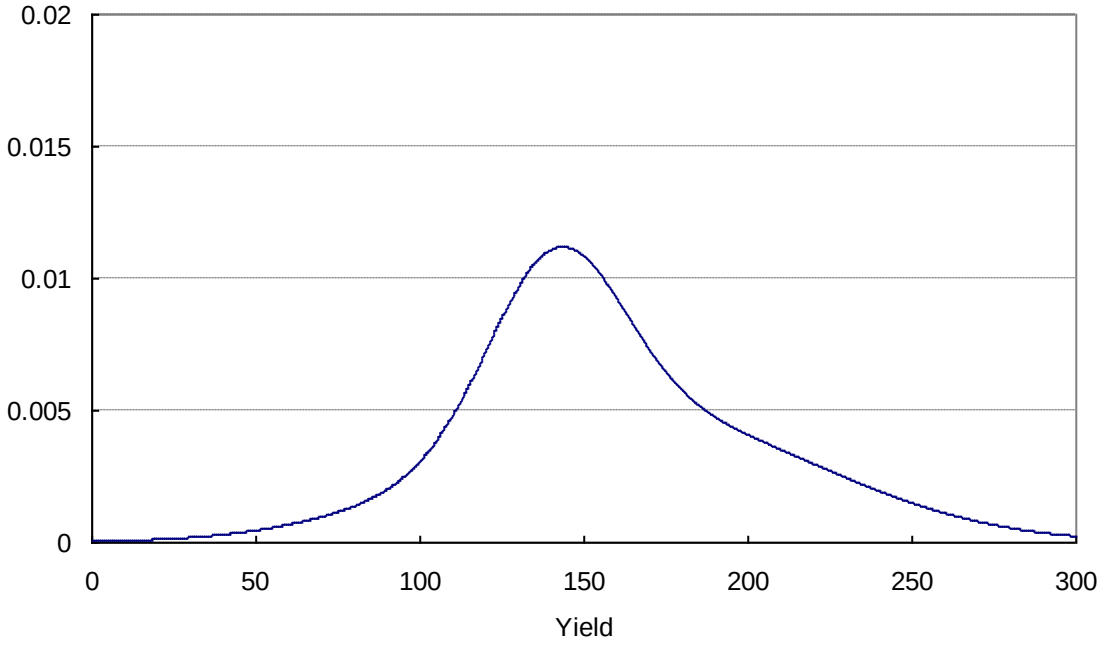


Figure 10: Standard Deviation Parameters Recovered: All Illinois Counties

Figure 11 illustrates the estimated conditional yield distributions for two given counties. Greene County represents the counties that have positively skewed distributions recovered, while Alexander County represents the opposite. A closer look at the counties which recover positively skewed distribution reveals that 28 of them are located in southern Illinois. Though Illinois lies entirely in the Interior Plains, it has three major geographical divisions—Northern Illinois, Central Illinois, and Southern Illinois (Figure 12). Southern Illinois, also known as "Little Egypt", is located near the juncture of the Mississippi River and Ohio River. This region can be distinguished from the other two by



(a)



(b)

Figure 11: Estimated Conditional Yield Distribution: (a) Greene County, Illinois; (b) Alexander County, Illinois

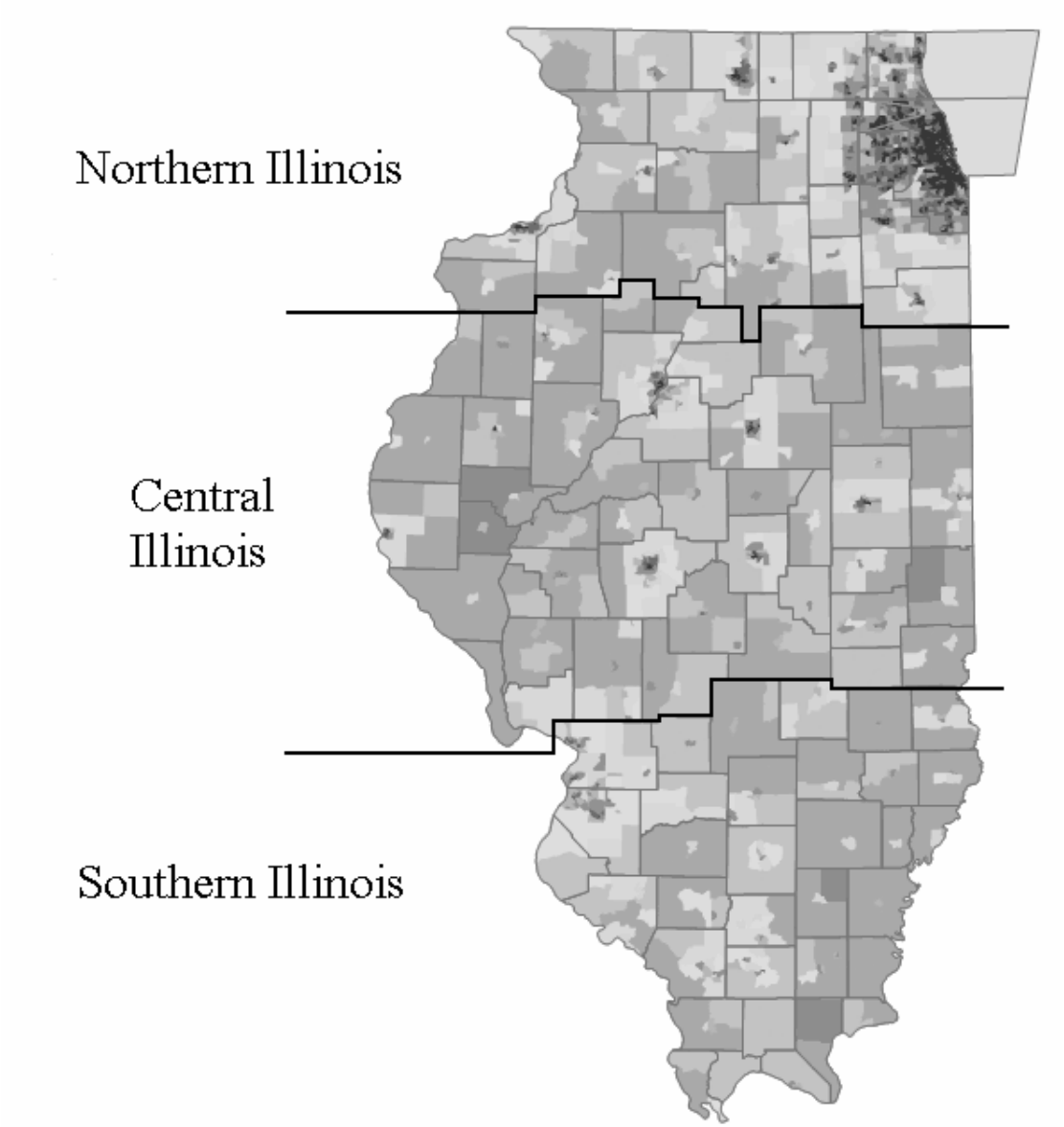


Figure 12: Three Major Geographical Divisions of Illinois

its warmer climate, different mix of crops, more rugged topography.[9] Most importantly, Southern Illinois has lower yield level on average over the years. Table 3 compares average county yield between 68 counties in Northern and Central Illinois and 34 counties in Southern Illinois. The explanation for the unexpected positive skewness lies in the significant yield difference between the two regions.

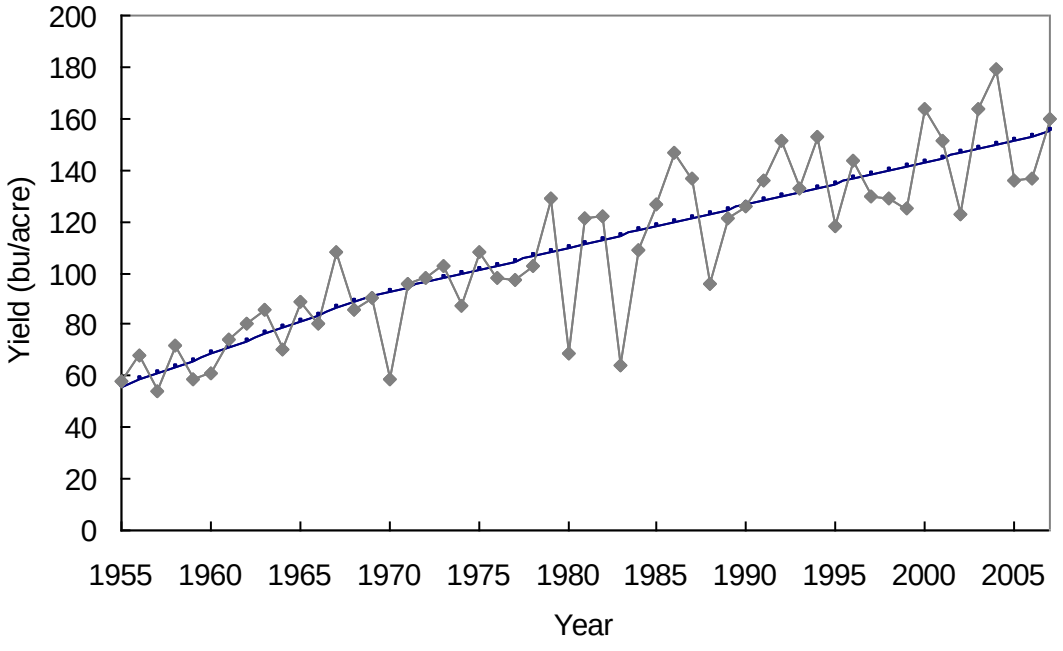
[9] http://en.wikipedia.org/wiki/Illinois (June 2009)

Table 3: Average County Yield Comparison within Illinois

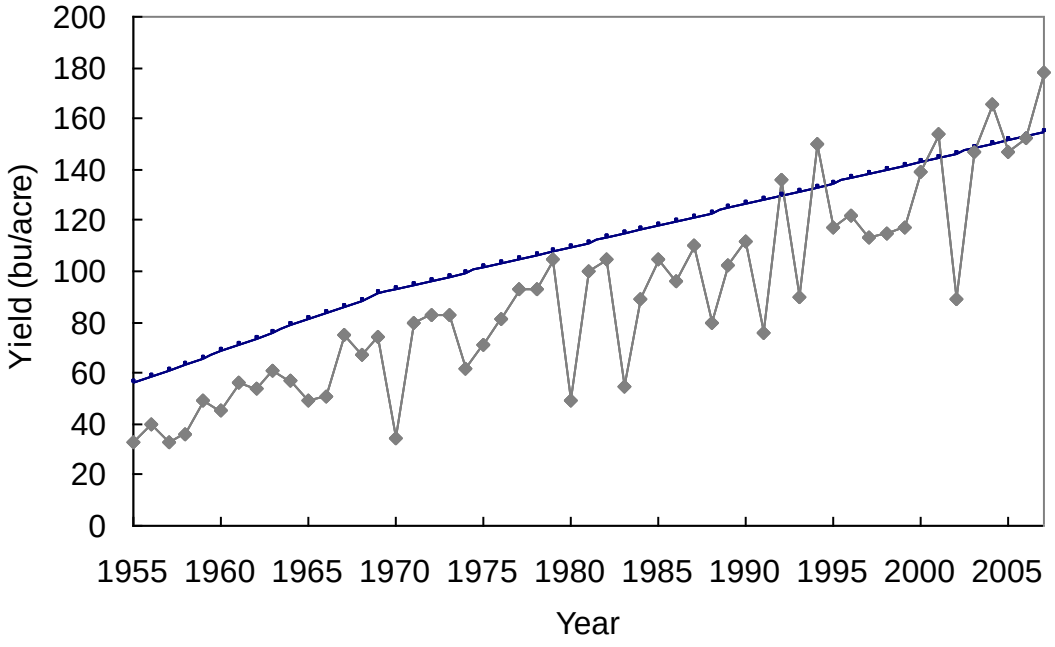
	1960	1970	1980	1990	2000	1955-2007
Northern and Central Illinois	68.8 (8.35)	77.2 (9.82)	97.7 (15.52)	129.9 (8.16)	152.1 (14.44)	114.9 (33.86)
Southern Illinois	49.6 (5.80)	42.6 (9.18)	60.3 (14.36)	98.0 (11.40)	134.2 (13.07)	87.3 (31.46)

* Values in brackets are standard deviations

Remember that the parameters indexing the spline model are assumed not to vary among counties, so all counties have the same estimate of central tendency of yield. However, counties in Southern Illinois have much lower yields on average over years, thus the central tendency estimated by all 102 counties tends to overestimate the yields of the 34 counties in the south. Figure 13 illustrates the estimated central tendency of yield and the realized yields for the same two counties in Figure 11. We can see clearly that the



(a)



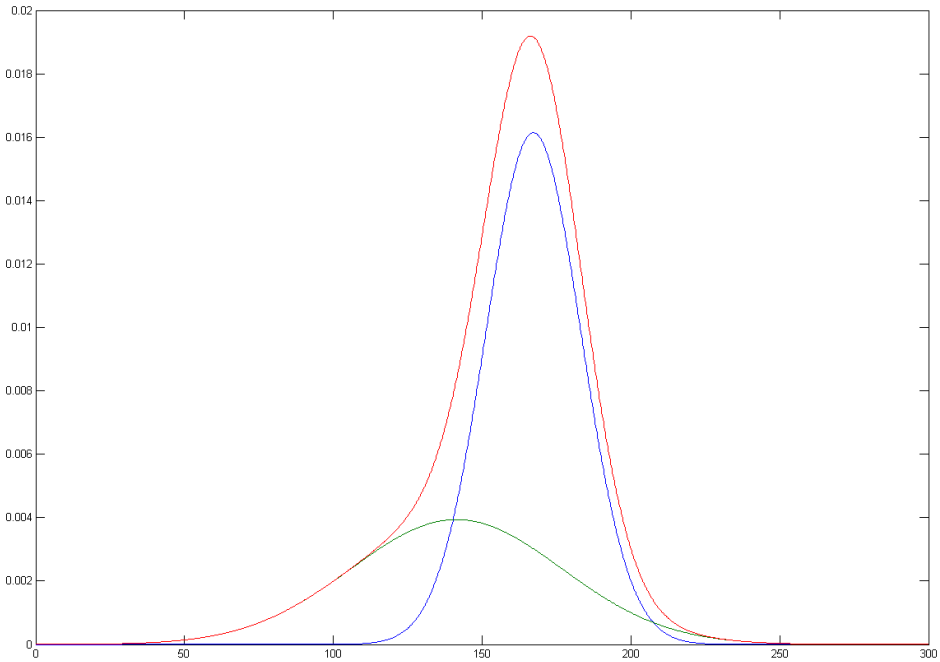
(b)

Figure 13: Estimated Central Tendency and Realized Yields: (a) Greene County, Illinois; (b) Alexander County, Illinois

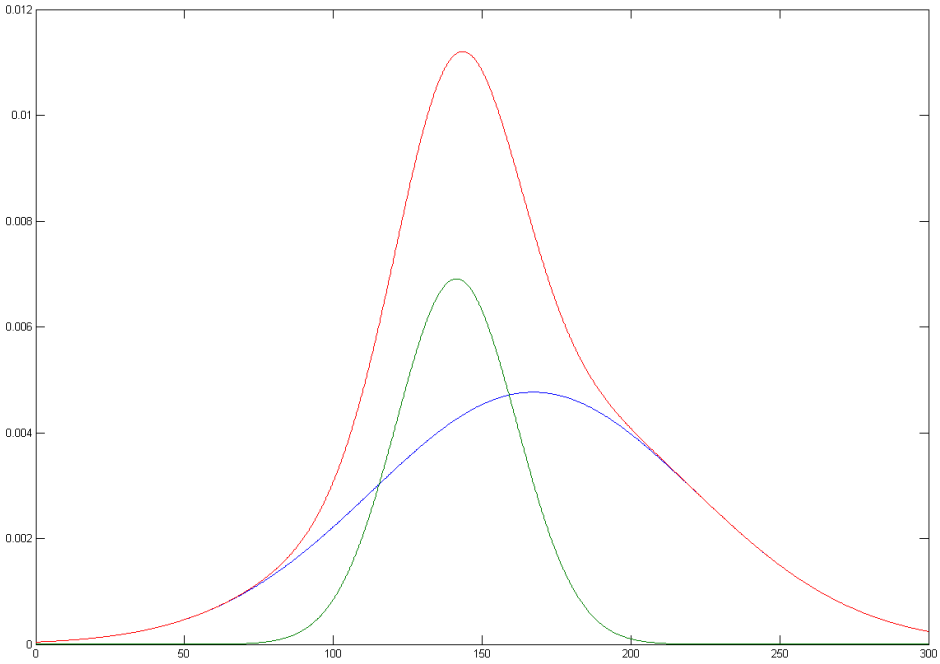
central tendency fit the Greene County yield much better than that of Alexander County where the central tendency overestimates yields in most years. In light of this systematic overestimation, the mean of the fitted residuals is far from zero.[10] Instead, most of the fitted residuals are deeply negative. At the same time, the location parameters and the mixture parameter of the two normal distributions are fixed among all counties and only the two variance parameters can differ by county. Intuitively, deep negative fitted residuals correspond to years with severe catastrophes, which by assumption are realizations from the secondary normal distribution. When only several deep negative fitted residuals occur, increase of the variance of the secondary normal distribution can

[10] Recall that the mean of the disturbance distribution is restricted to be zero in order to recover the intercept parameters.

increase the likelihood thus accommodate those deep negative residuals (Figure 14a). However, for counties that recover positively skewed distributions, there are few positive fitted residuals but negative fitted residuals both big in magnitude and large in quantity. In this case, solely increasing the variance of the secondary normal distribution cannot accommodate them and maximize the likelihood. In other words, they are more likely to come from the major normal distribution. So the variance parameter of the major normal distribution is increased to maximize the likelihood. Thus, positively skewed distribution is recovered (Figure 14b). In fact, the assumption that the secondary normal distribution captures catastrophic years is violated in these counties, since most years are "catastrophic" due to overestimation. Now we can conclude that the unexpected positive skewness of 29 estimated yield distributions are artifactual as a result of systematic overestimation of the central tendencies.



(a)



(b)

Figure 14: A Mixture of Two Normal Distributions Recovered: (a) Greene County, Illinois; (b) Alexander County, Illinois

Now, given the estimation results and what we found above, recall Figure 8. The separation of the estimated joint density of the random effects is more likely to be the result of the inflexible model specification, rather than the omission of some county specific attributes.

5.2 Premium Rates

Recall, premium rates are expressed as expected indemnity as a percentage of total liability. For a contract with coverage level of λ% of the expected yield y^e, the premium rate is given as

(23) Premium Rate (%) =

$$\frac{\text{Expected Indemnity}}{\text{Liabilty}} = \frac{\Pr(Y < \lambda y^e)[\lambda y^e - \mathrm{E}(Y \mid y < \lambda y^e)]}{\lambda y^e}$$

where the expectation operator and probability measure are taken with respect to the conditional yield density. Table 4 presents summary statistics of the SNP rates and the empirical rates for Illinois all practices corn.

Table 4: Summary Statistics of GRP Premium Rates: Illinois Corn—All Practices

Method	Coverage Level				
	70%	75%	80%	85%	90%
SNP Rates					
mean	1.8654	2.4331	3.1495	4.0439	5.1530
std. deviation	0.8970	1.0335	1.2078	1.4348	1.7136
minimum	0.6744	1.0719	1.6419	2.3841	3.2855
maximum	4.1405	5.0112	6.1020	7.4637	9.0959
Empirical Rates					
mean	1.7242	2.4152	3.2931	4.3598	5.6274
std. deviation	2.3206	3.1227	4.0711	5.1170	6.2485
minimum	0.0000	0.0000	0.0000	0.0916	0.2962
maximum	9.4218	12.1942	15.3938	19.2574	23.1982

Empirical premium rates are calculated as sum of all percentage shortfalls divided by the number of years. In fact, it is the simplest nonparametric approach available for making inferences about distributions. However, large samples are needed to accurately calculate rates because no smoothing is undertaken and the distributions are not required to be continuous (Goodwin and Ker, 1998). The SNP rates are calculated numerically from the recovered yield distributions of each county. The SNP rates are higher than empirical rates at 70% and 75% levels, while lower than empirical rates at coverage levels above 80%. Given the small number of realizations, empirical rates tend to put

inadequate mass in the extreme lower tails of the estimated density, as opposed to SNP rates. In addition, empirical rates have much higher standard deviations at all coverage levels, suggesting that rates are more volatile among counties: for counties having experienced few extreme low yields, the rates can be zero for 80% coverage level. In contrast, SNP approach, which yields more stable rates, considers the data from all counties when estimating the tail probabilities for any given county.

5.3 Concluding Remarks

The main objective of this thesis is to improve the accuracy of premium rates, thereby improving the efficiency of the GRP programs. Accurate premiums rates require proper estimation of the conditional yield distributions, including accurately predicting mean yield and accurately estimating the mass about the predicted mean yield. A spline model with one knot point is used to capture the central tendency and to predict mean yield. A flexible mixture of two normal distributions is used to represent the disturbance distribution. Parameters indexing the mixture distribution follow a mixed effects regression model where the joint density of the random effects is estimated by a Hermite expansion. Seminonparametric maximum likelihood estimation technique is employed to estimate the spline model and disturbance distribution simultaneously. County specific yield distributions are attained from the posterior modes of the random effects and estimates of the fixed effects.

SNP maximum likelihood approach makes more efficient use of the data in that yield correlation among counties is explicitly modeled. As a result, SNP rates are more consistent among counties. Also, SNP method tends to put relatively more mass in the

extreme lower tail of the estimated density, which helps make the premium rates at low coverage levels more reasonable, especially given the limited number of realizations for each county. It is shown in the thesis how conditional yield distributions depart from normality and that a mixture of two normal distributions could be used to represent conditional yield densities. The thesis also demonstrates that yield distributions do not necessarily belong to a parametric family of distributions that have been used in the past (e.g. beta, gamma).

It is discussed in Chapter 3 that some researchers have justified the one-knot linear spline temporal process on the following basis: while technology will directly increases average yields, the rate of farm consolidation will also increase average yields. Farm consolidation is greatest in the decades of the 1950s and 1960s and slow in the 1970s thus suggesting a one-knot spline. The knot point selected by likelihood value in this thesis is 1969, which can be considered as a support of the claim.

There are at least three areas requiring further study. First, the parameters indexing the spline model are assumed to be constant among counties in this thesis. Although this assumption helps limit the dimension of the estimator and achieve a gain in degrees of freedom, it also sacrifices the flexibility of the model to accommodate characteristics of different counties. For example, the estimated central tendency does not fit the counties from Southern Illinois as well as the other counties. The inflexibility of the central tendency even induces artifactual positive skewness in estimated yield distributions. More flexible temporal process like AR(p) model might be a better choice for the central tendency.

Second, the correlation structure among counties is modeled with an inadequate

set of county specific attributes. The multimodality found in the joint density of random effects might be due to some omitted influential county specific attributes. If this is the case, the model will be substantially improved if the omitted factor is identified and included in the inter-county regression functions. Another possible explanation for the multimodality found in the joint density of random effects is the systematic overestimation of a group of counties. If this is the case, then a more flexible temporal process like AR(p) may remove the multimodality without adding any more county specific attributes.

Last but not least, further research is needed to compare competitive estimation methods and analyze the economic implications of the estimation errors. A simulation performed in Ker and Coble (2003) exemplifies a good way of doing this. In the simulation, competing estimators are used to estimate a set of yield densities and derive the associated premium rates, including the rating methodology of Risk Management Agency (RMA). Then they evaluate the competing estimators by calculating out-of-sample loss ratios based on decision rules for retaining or ceding GRP crop insurance contracts. This simulation is appealing from an economic and policy perspective because private insurance companies are allowed to retain or cede, *ex ante* and subject to constrains, varying portions of the realized underwriting gains/losses of every federally subsidized crop insurance contract it sells. Similarly, such a simulation will enhance our understanding of the SNP estimators.

APPENDIX: ADDITIONAL FORTRAN CODE FOR NLMIX

```
      subroutine density(isw)
      implicit real*8 (a-h,o-z)
      implicit integer*4 (i-n)
      save

      include "global.f"

      real*8 yld,x1,x2,macres,sacres
      real*8 data
      real*8 sig1,sig2,expz1,expz2,vpow
      real*8 P,Pwsig1,Pwsig2,PwO,PwOx1,PwOx2,Pw12,PwO2
      real*8 O,O2,S,S2
      real*8 Plag,Pws1lag,Pws2lag,PwO1lag,PwOx1lag,PwOx2lag,Pw12lag,
                PwO2lag
      real*8 f,fwO,fwS,f1,f2,fwO2,fwS2

      integer*4 i,j,isw
      integer*4 id0,id0lag
      integer*4 count,point,iptr

      character*15 z1out,z2out

      integer*4 mrows,mcols,debug

      parameter (mrows=6,mcols=5406,debug=1)
      parameter (vpow=2.d0)

      dimension count(mcols),point(mcols)
      dimension data(mrows,mcols)


*     If isw=1, read and setup local data; compute nobs.

      if (isw.eq.1) then

       if (iunit9.eq.1) then
         j=0
         do i=1,12
           if (outfil(i:i).eq.' ') then
             j=i
             go to 5
           end if
         end do
5         continue
```

```
      if (j.eq.0) then
        z1out=outfil//'.z1'
        z2out=outfil//'.z2'
      else
        z1out=outfil(1:j-1)//'.z1'
        z2out=outfil(1:j-1)//'.z2'
      end if
      open (unit=21,file=z2out,status='unknown',form='formatted')
      open (unit=22,file=z1out,status='unknown',form='formatted')
    end if

    open (unit=14,file='s1969.dat',status='old',form='formatted')

    nobs=0
    iptr=1
    count(iptr)=0
    point(iptr)=1
    id0lag=1

    do i=1,mcols
      read(14,14001,end=10)
   &     id0,yld,x1,x2,macres,sacres

      data(1,i)=yld
      data(2,i)=x1
      data(3,i)=x2
      data(4,i)=macres
      data(5,i)=sacres

      nobs=nobs+1

      if (id0lag.eq.id0) then
        count(iptr)=count(iptr)+1
      else
        iptr=iptr+1
        count(iptr)=1
        id0lag=id0
        point(iptr)=i
      end if
    end do
10    continue

    if (iptr.ne.ncases) then
      write(3,3001)
      stop
    end if
```

```
      close(14)
      return
     end if

*    If isw=2, compute the density.

     if (isw.eq.2) then
      macres=data(4,point(id))
      sacres=data(5,point(id))

      expz1=z(1)
      expz2=z(2)

      sig1=dexp(tau(1)+macres*tau(2)+expz1)
      sig2=dexp(tau(7)+macres*tau(8)+expz2)

      Plag = 1.d0
      Pws1lag = 0.d0
      PwO1lag = 0.d0
      PwOx1lag = 0.d0
      PwOx2lag = 0.d0
      Pws2lag = 0.d0
      PwO2lag = 0.d0
      Pw12lag = 0.d0

      do i=point(id),point(id)-1+count(id)
       yld=data(1,i)
       x1=data(2,i)
       x2=data(3,i)
       O = tau(3)+tau(4)*x1+tau(5)*x2
       S = sig1
       call normalh(yld,O,S,vpow,f1,fwO,fwS)
       O2 = tau(6)+tau(4)*x1+tau(5)*x2
       S2 = sig2
       call normalh(yld,O2,S2,vpow,f2,fwO2,fwS2)
       f = tau(9)*f1+(1-tau(9))*f2
       P = Plag*f
       Pwsig1 = Plag*fwS*tau(9) + Pws1lag*f
       Pwsig2 = Plag*fwS2*(1-tau(9)) + Pws2lag*f
       PwO = Plag*tau(9)*fwO + PwO1lag*f
       PwOx1 = Plag*x1*(fwO*tau(9)+fwO2*(1-tau(9)))+PwOx1lag*f
       PwOx2 = Plag*x2*(fwO*tau(9)+fwO2*(1-tau(9)))+PwOx2lag*f
       PwO2 = Plag*(1-tau(9))*fwO2 + PwO2lag*f
       Pw12 = Plag*(f1-f2)+Pw12lag*f
```

```
        Plag = P
        Pws1lag = Pwsig1
        Pws2lag = Pwsig2
        PwO1lag = PwO
        PwO2lag = PwO2
        PwOx1lag = PwOx1
        PwOx2lag = PwOx2
        Pw12lag = Pw12
      end do

      d = P
      dwtau(1)=Pwsig1*sig1
      dwtau(2)=Pwsig1*macres*sig1

      dwtau(3)=PwO
      dwtau(4)=PwOx1
      dwtau(5)=PwOx2
      dwtau(6)=PwO2

      dwtau(7)=Pwsig2*sig2
      dwtau(8)=Pwsig2*macres*sig2


      dwtau(9)=Pw12

      dwz(1)=Pwsig1*sig1
      dwz(2)=Pwsig2*sig2

      return
    end if

*     If isw=3, write individual effects to unit 9.

    if (isw.eq.3) then
      write(9,9001) id,sig1,sig2
      write(21,21000) z(1)
      write(22,22000) z(2)
    end if

*     If isw=4, print some debugging information.

    if (isw.eq.4) then

      if (iunit9.eq.1) then
        close(21)
        close(22)
```

```
       end if

       if ((iprint.eq.0).and.(debug.ne.0)) then
         do i=1,ncases
           write(3,3008) i,point(i),count(i)
           write(3,3009) (data(j,point(i)),j=1,mrows)
           write(3,3009) (data(j,point(i)-1+count(i)),j=1,mrows)
         end do
         write(3,3002) sig1,sig2
         write(3,3005) Pws1lag,Plag
         write(3,3006) P,Pwsig1
         write(3,3007) O,S,f,fwO,fwS
         return
       end if
     end if

     return
3001  format(1x,'hao,error, density, error reading data')
3002  format(1x,'sig1,sig2'/1x,2f20.8)
3005  format(1x,'Pws1lag,Plag'/1x,2f20.8)
3006  format(1x,'P,Pwsig1'/1x,2f12.8)
3007  format(1x,'O,S,f,fwO,fwS'/1x,5f11.7)
3008  format(1x,'i,point,count',1x,3i6)
3009  format(1x,'data',1x,18f8.4)
9001  format(i6,3d30.18)
14001 format(i4,f7.1,2f8.1,2f9.4)
21000 format(1x,f28.20)
22000 format(1x,f28.20)
23000 format(1x,f28.20)
     end
```

REFERENCES

Atwood, J., S. Shaik, and M. Watts. "Are Crop Yields Normally Distributed? A Reexamination." *Amer. J. Agr. Econ.*, 85(4) (November 2003): 888–901.

Babcock, B.A. and D.A. Hennessy. "Input Demand under Yield and Revenue Insurance." *Amer. J. Agr. Econ.*, 78(2) (May 1996): 416-427.

Bessler, D.A. "Aggregated Personalistic Beliefs on Yields of Selected Crops Estimated Using ARIMA Process." *Amer. J. Agr. Econ.*, 62 (November 1980): 666-674.

Botts, R.R. and J.N. Boles. "Use of Normal-Curve Theory in Crop Insurance Ratemaking." *Journal of Farm Economics*, 40(3) (August 1958): 733-740.

Breusch, T. and A. Pagan. "A Simple Test for Heteroscedasticity and Random Coefficient Variation." *Econometrica*, 47 (1979): 1287-1294.

Coble, K.H., T.O. Knight, R.D. Pope, and J.R. Williams. "Modeling Farm-Level Crop Insurance Demand with Panel Data." *Amer. J. Agr. Econ.*, 78(2) (May 1996): 439-447.

Coble, K.H. and T.O. Knight. "Crop Insurance as a Tool for Price and Yield Risk Management." *A Comprehensive Assessment of the Role of Risk in U.S. Agriculture*, R.E. Just and R.D. Pope, Massachusetts: Kluwer Academic Publishers, 2002.

Davidian, M. and A.R. Gallant. "Smooth Nonparametric Maximum Likelihood Estimation for Population Pharmacokinetics with Application to Quinidine." *Journal of Pharmacokinetics and Biopharmaceutics*, 20(5) (1992): 529-556.

Davidian, M. and A.R. Gallant. "The Nonlinear Mixed Effects Model with a Smooth Random Effects Density." *Biometrika*, 80(3) (September 1993): 475-488.

Deng, X., B.J. Barnett, and D.V. Vedenov. "Is There a Viable Market for Area-Based Crop Insurance?" *Amer. J. Agr. Econ.*, 89(2) (May 2007): 508–519.

Gallagher, P. "U.S. Soybean Yields: Estimation and Forecasting with Nonsymmetric Disturbances." *Amer. J. Agr. Econ.*, 69(4) (November 1987): 796-803.

Gallant, A.R. and D.W. Nychka. "Semi-Nonparametric Maximum Likelihood Estimation." *Econometrica*, 55(2) (March 1987): 363-390.

Glauber, J.W. and K.J. Collins. "Risk Management and the Role of the Federal Government." *A Comprehensive Assessment of the Role of Risk in U.S. Agriculture*, R.E. Just and R.D. Pope, Massachusetts: Kluwer Academic Publishers, 2002.

Goodwin, B.K. and A.P. Ker. "Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts." *Amer. J. Agr. Econ.*, 80(1) (February 1998): 139-153.

Goodwin, B.K. and A.P. Ker. "Modeling Price and Yield Risk." *A Comprehensive Assessment of the Role of Risk in U.S. Agriculture*, R.E. Just and R.D. Pope, Massachusetts: Kluwer Academic Publishers, 2002.

Greene, W.H. *Econometric Analysis*, 4th Edition, New Jersey: Prentice-Hall, 2000.

Hennessy, D.A. "Crop Yield Skewness and the Normal Distribution." *Journal of Agricultural and Resource Economics*, 34(1) (April 2009): 34-52.

Just, R.E. and Q. Weninger. "Are Crop Yields Normally Distributed?" *Amer. J. Agr. Econ.*, 81(2) (May 1999): 287-304.

Ker, A.P. "Rating and Yield Predicting Procedures for the Group Risk Federal Crop Insurance Program: A Nonparametric Approach." Ph.D. Dissertation, North Carolina State University (1996).

Ker, A.P. and B.K. Goodwin. "Nonparametric Estimation of Crop Insurance Rates Revisited." *Amer. J. Agr. Econ.*, 82(2) (May 2000): 463-478.

Ker, A.P. and K. Coble. "Modeling Conditional Yield Densities." *Amer. J. Agr. Econ.*, 85(2) (May 2003): 291-304.

Koenker, R. and G. Bassett. "Robust Tests for Heteroscedasticity Based on Regression Quantiles." *Econometrica*, 50 (1982): 43-61.

Moss, C.B. and J.S. Shonkwiler. "Estimating Yield Distributions with a Stochastic Trend and Nonnormal Errors." *Amer. J. Agr. Econ.*, 75(4) (November 1993): 1056-1062.

Nadolnyak, D., D. Vedenov, and J. Novak. "Information Value of Climate-Based Yield Forecasts in Selecting Optimal Crop Insurance Coverage." *Amer. J. Agr. Econ.*, 90(5) (2008): 1248–1255.

Nelson, C.H. and P.V. Preckel. "The Conditional Beta Distribution as a Stochastic Production Function." *Amer. J. Agr. Econ.*, 71(2) (May 1989): 370-378.

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

Norwood, B., M.C. Roberts, and J.L. Lusk. "Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions." *Amer. J. Agr. Econ.*, 86(4) (November 2004): 1032–1043.

Ozaki, V.A., S.K. Ghosh, B.K. Goodwin, and R. Shirota. "Spatio-Temporal Modeling of Agricultural Yield Data with an Application to Pricing Crop Insurance Contracts." *Amer. J. Agr. Econ.*, 90(4) (November 2008): 951–961.

Racine, J. and A. Ker. "Rating Crop Insurance Policies with Efficient Nonparametric Estimators that Admit Mixed Data Types." *Journal of Agricultural and Resource Economics*, 31(1) (April 2006): 27-39.

Ramirez, O.A. "Estimation and Use of a Multivariate Parametric Model for Simulating Heteroskedastic, Correlated, Nonnormal Random Variables: The Case of Corn Belt Corn, Soybean, and Wheat Yields." *Amer. J. Agr. Econ.*, 79 (February 1997): 191–205.

Ramirez, O.A., S. Misra, and J. Field. "Crop-Yield Distributions Revisited." *Amer. J. Agr. Econ.*, 85(1) (February 2003): 108–120.

Sherrick, B.J., F.C. Zanini, G.D. Schnitkey, and S.H. Irwin. "Crop Insurance Valuation under Alternative Yield Distributions." *Amer. J. Agr. Econ.*, 86(2) (May 2004): 406–419.

Skees, J.R., J.R. Black, and B.J. Barnett. "Designing and Rating an Area Yield Crop Insurance Contract." *Amer. J. Agr. Econ.*, 79(2) (May 1997): 430-438.

Taylor, C.R. "Two Practical Procedures for Estimating Multivariate Nonnormal Probability Density Functions." *Amer. J. Agr. Econ.*, 72(1) (February 1990): 210-217.

U.S. Department of Agriculture. *FY 2009 Budget Summary and Annual Performance Plan*, 2008.

U.S. Government Accountability Office. "Crop Insurance: Continuing Efforts Are Needed to Improve Program Integrity and Ensure Program Costs Are Reasonable." GAO-07-944T, Washington, D.C., June 7, 2007.