2010

Stochastic trends in crop yield density estimation

David I. Maradiaga
Louisiana State University and Agricultural and Mechanical College, dmarad1@tigers.lsu.edu

STOCHASTIC TRENDS IN CROP YIELD DENSITY ESTIMATION

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science

in

The Department of Agricultural Economics and Agribusiness

by
David I. Maradiaga
B.S., Escuela Agrícola Panamericana, El Zamorano, 2005
August 2010

ACKNOWLEDGMENTS

I would like to show my sincere gratitude to God, who has provided me happiness, love, immeasurable blessings, determination and strength that helped me to finish this journey.

I would like to thank Dr. William B. Richardson and the LSU-AgCenter for financially assisting my studies, and for their efforts in the agricultural development of Honduras.

I would like to express the deepest appreciation to my committee chair, Dr. Hector O. Zapata, for his mentorship and trust that I could take challenges and develop a critical thinking ability. My gratitude is also extended to the members of my graduate committee: Dr. Jeffrey M. Gillespie and Dr. Michael E. Salassi, for their valuable reviews and suggestions; and to Dr. Gail L. Cramer for his advice and encouragement throughout the Master of Science program.

I would like to show profound admiration to my *fiancée*, Aude Liliana Pujula, for being such an exceptional person, as well as a great colleague and research team member. Your love enhances my atmosphere and your work inspired me to pursue higher objectives.

I would like to show my gratitude to Alejandro Maradiaga and Orbelina Pineda (parents) and to Johana and Alejandra (sisters), for their love, advice, encouragement, and huge efforts made for my education. My appreciation is also extended to Antoine-Robert Pujula (*petit beau-père*), Anne-Marie Pujula (*petite belle-mère*) and Jeanne Carrere (*mami*), for their love, advice and kindness.

I would like to thank Mrs. Maureen Hewitt, Stephanie Gil, and Michelle Rzepka for their valuable comments in this manuscript. My gratitude is also extended to all my friends in the AgEcon Dept. and Zamorano Agricultural Society at LSU for their hard work and good times.

I would like to thank Alicia Ryan for her hospitality and consideration during my time here in Baton Rouge.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACRE	Average Crop Revenue Election
ADF	Augmented Dickey Fuller
AIC	Akaike's Information Criterion
ARIMA	AutoRegressive Integrated Moving Average
ARMA	AutoRegressive Moving Average
Bu	Bushel
CL	Coverage Level
DGP	Data Generating Process
ECID	Error Component Implicit Detrending
EDF	Empirical Density Function
FCIC	Federal Crop Insurance Corporation
GRP	Group Risk Plan
I (1)	Integrated of order one
IHST	Hyperbolic Sine Transformation
MA	Moving Average
NASS	National Agriculture Statistics Service
OSLLF	Out-of-Sample-Log-Likelihood Function
PC	Correct Probability (Percentile)
PDF	Probability Density Function
PE	Percent Error
PI	Incorrect Probability (Percentile)
RMA	Risk Management Agency
RW	Random Walk
RWD	Random Walk with Drift
RWDT	Random Walk with Drift and Trend
ST	Stationary
TS	Trend Stationary
TY	Trigger Yield
USDA	United States Department of Agriculture

ABSTRACT

The search for improved methods of estimating crop yield density functions has been a theme of recurrent research interest in agricultural economics. Crop yield density functions are the statistical instrument that generates probability estimates of yield risk, and risk is an important decision variable in production agriculture. Recent research in crop yield density estimation suggests that yield probability estimates can be sensitive to the way yield data are filtered, and if true, then the search for an "adequate filter" is warranted. Such a quest is pursued in this study. It is proposed that unit-root tests can be used to identify the time-series properties of yields and that the outcome of these tests makes the choice of an appropriate filter trivial. Once a filter has been chosen, then nonparametric methods can be used to more flexibly fit a crop yield density function.

The study uses state and county level (aggregated) yield data for corn and soybeans in Arkansas and Louisiana for the period 1960-2008, comprising 121 yield series. The results identify three main types of yield processes (and filters), namely, a unit-root (first differences), a trend stationary process (detrending), and stationary (remove the mean). More specifically, the study finds that for Louisiana soybeans, for example, 73% of the county yields studied can be represented by a unit-root process, 12% followed a trend stationary process, and the remaining 15% were stationary. One important implication of this finding is that the use of a universal yield filter may generate inaccurate yield probability estimates, which translates into inaccurate estimates of crop insurance risk *premia*. To shed light into relevance of these findings, yield density functions were estimated under alternative filtering scenarios and pairwise probability estimates compared. In particular, the results suggest sizeable differences in the two estimates, which at times can reach -1,153.65%. In addition to providing a detailed analysis of the findings,

the study assessed the relevance of the findings in the context of two current risk management programs, namely a group risk plan (GRP) and average crop revenue election (ACRE) program. Limitations of the study are also highlighted.

1. INTRODUCTION

About half a century ago, the *Journal of Farm Economics* published a paper on the probability distribution of field crop yields that established an empirical foundation for agricultural decision making under risk (Day, 1965). Day stated that one of the primary purposes for seeking a better understanding of the stochastic properties of field crop yields is the conversion of uncertainty into risk, and that “until a science of yield probabilities can be developed, correct decisions in agriculture are virtually impossible” (Day, 1965, p. 714). This statement is still true today and has implications for the estimation of crop insurance *premia* and other very important risk analyses in agriculture.

In reviewing the vast and continual literature on the subject, it is found that the science of density estimation has improved, though the search for statistical reliability continues. This search is documented extensively in the statistical literature and applications to crop yield density estimation using historical (time-series) data abound. Initial research on the subject focused on the use of parametric methods (*e.g.,* Botts and Boles, 1958; Day, 1965; Gallagher, 1986; 1987; Nelson and Preckel, 1989; Nelson, 1990 and Atwood Shaik and Watts, 2003), but recently there is a considerable convergence of literature towards nonparametric methods (*e.g.,* Goodwin and Ker, 1998; Turvey and Zhao, 1999; Ker and Coble 2003 and Norwood, Roberts and Lusk, 2004).

Literature in crop yield density estimation also has recorded the use of different data types such as farm-level (*e.g.,* Sherick *et al.,* 2004), experimental (*e.g.,* Day, 1965), aggregated (*e.g.,* Gallagher, 1986), cross sectional (*e.g.,* Kaylen and Koroma, 1991), panel (*e.g.,* Goodwin and Ker, 1998) and time-series (*e.g.,* Moss and Shonkwiler, 1993).

The estimation of crop yield density functions requires random (nonsystematic) data. Actual yield data, however, contain trends, autoregressive and moving average terms, and other effects that need to be removed prior to fitting a density. To remove these systematic effects, a transformation (*e.g.,* detrending and differencing) is usually needed in order to generate a random process. [1] Although density estimation techniques have flourished, less emphasis has been given to the appropriateness of the data transformations and how such transformations can impact the reliability of distribution functions and the resulting probability estimates using time-series data.

In the agricultural economics literature, the upward trend of crop yields is often represented by a linear trend suggesting that crop yields are increasing at a constant rate from year to year. Once the trend line is estimated, it is assumed that this kind of pattern continues indefinitely and can be used for forecasting. For that reason, such a model is called a "deterministic trend." For up-trending crop yields, this trend can be visualized as a straight line where the slope on a trend variable (the horizontal axis on Figure 1.1) represents the magnitude of the increase in yields per year. The simplicity of this model in risk analyses is appealing. In crop yield density estimation the residuals (actual yields minus the estimated trend line) are used to identify the proper empirical density function (EDF), and this EDF is used for risk calculations.

While some crop yields can follow a deterministic model, existing empirical evidence with aggregated time-series suggests that other, perhaps more complex processes, may be appropriate. Crop yields may change due to the influence of location specific factors such as soil types and weather, the rate of technology adoption can also create a geographic variation, and

[1] A random process is defined as a stochastic process that has a constant mean and variance, and whose covariance between two elements of the sequence only depends on the length of time (Hamilton, 1994).

there also are farm specific factors impacting yields. The combined effect of such changing factors may result in yields similar to those shown in Figure 1.1. For instance, it is evident that a severe drought during the 1930s (Dust Bowl era) had effects on yields that can be seen as shocks introduced into the series, while the uneven adoption of technology (*e.g.*, use of nitrogen) created a stronger trend but also less predictable since the 1950s. Consequently, average yields have fluctuated from decade to decade, as have their related variances.

As is well known from the time-series literature, a process whose mean and variance are changing is called "nonstationary," one example of which is a random walk. At the expense of oversimplification, a random walk can be thought of as a process with trends whose slope and direction change in an unpredictable manner. Such a process is also called a "stochastic trend" and can be visualized for upward trending crop yields as the irregular uptrending series represented by the dark blue line in Figure 1.1. If series have a stochastic trend, then it is common practice to remove the "trends" using first-differences or detrending. This process of transforming yields to a random process suitable for density estimation is often called filtering.

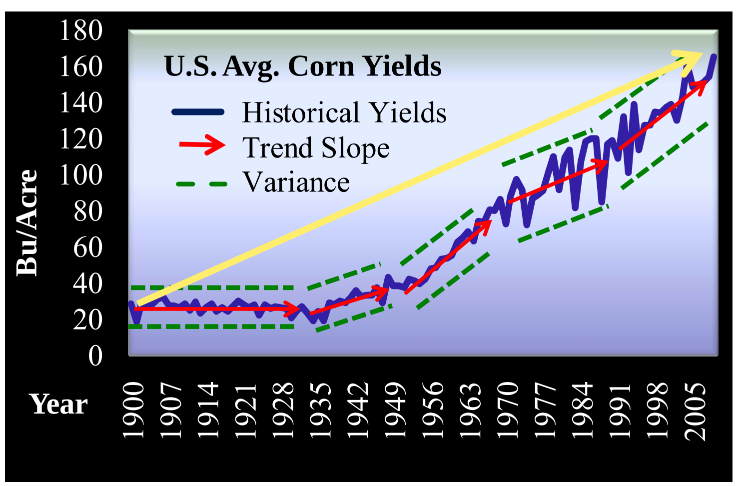

Figure 1.1. US Historical Corn Yields.

1.1 Problem Statement

In agricultural economics and in particular in the crop yield risk literature, it is common to filter the data prior to density estimation to account for trends and systematic effects. Enders (1995) stated that the assumption that an upward trend of a series can be represented by a linear time trend is controversial which raises the following question: Does filtering (transforming) historical crop yields prior to density estimation impact yield probability estimates?

An early study by Zapata and Rambaldi (1989) found that an arbitrary transformation to crop yields can generate series with properties than are much different from those of the underlying stochastic process. They also suggested that unit-root tests constitute a useful tool to determine which transformation to employ. In the context of crop yield density estimation, the general question above can be broken down in two. First, can unit-root test be used to identify appropriate filters for historical yield data? Second, are the probability estimates sensitive to the choice of filters?

1.2 Justification

As implied by Day (1965), a better science of yield probabilities can result in improved decision making in agriculture. There is still a limited amount of literature on the evaluation of the different data transformation techniques. Crop yield density functions are the main statistical instrument used to calculate yield probabilities. Therefore, if the filtering step in density estimation can be improved, then the likelihood of estimating incorrect probabilities is diminished. Improved probability estimates are invaluable in various applications. In crop insurance, for example, risk *premia* are set based on estimates of yield risk. If yield risk is more correctly measured, then actuarial tables for crop insurance product could be better suited to the local yield risk experience.

Researchers have argued that the determination of an accurate measurement of crop yield risk is essential for crop insurance contracts and premium rating. However, if inappropriate filtering impacts the estimate of an empirical density function (EDF), then this may lead to a the wrong selection of the crop insurance protection level, unjustified crop insurance premium rates, and lower crop insurance participation rates. The empirical findings of this study should be of value to researchers who risk management education, to stakeholders of the U.S. crop insurance program, and to farmers who buy insurance products.

1.3 Objectives

The general purpose of this thesis is to identify and empirically test a method for choosing filters in crop yield density estimation with time-series data.

The specific objectives of this thesis are to:

1. Apply unit-root tests to historical crop yields and identify appropriate filters prior to density estimation.
2. Determine the impact of alternative filters on corn and soybean yield probability estimates.

1.4 Procedure

A proposition of this study is that alternative data filtering methods can lead to the identification of different density functions and probability estimates. A procedure capable of unmasking these possible differences is desired. Hence, the computation of a percent error between alternative probabilities is a plausible approach to reveal the hypothesized difference.

The empirical study uses historical corn and soybean yields (bu/acre) for Arkansas and Louisiana, obtained from the National Agricultural Statistics Service for the 1960-2008 period.

The data are aggregated county level yields for irrigated and non-irrigated crops for counties with a production history of at least 30 years and no more than five years of missing data. These data screening resulted in a total of 31 corn and soybeans producing counties in Arkansas and 25 corn and 34 soybeans producing parishes in Louisiana.

1.4.1 Objective One

Historical crop yields can be characterized as either trend deterministic or stochastic. The econometric literature on time-series provides numerous testing procedures, commonly referred to as unit-root tests, which can be used to identify the type of trend in yield series. More specifically, the augmented version of the Dickey-Fuller (1979) test (ADF) has been applied to Arkansas and Louisiana corn and soybean yields for the 1960-2008 period to identify the DGP. The test procedure is as follows (*e.g.,* Enders, 1995).

First, a regression of first differences on a constant, linear trend, lagged yield, and lagged first differences terms is performed. Subsequently, the t-statistic associated with the lagged yield coefficient is used to test the null hypothesis that the series has a unit-root at the 0.10 significance level.[2] If there is no unit-root, then it implies that the series is stationary. Otherwise, if there is a unit root, a joint test for the significance of the trend and a unit-root is performed. If the trend term is not significant, then the second step is executed. Otherwise, if there is a trend, then a unit-root test is carried out using a z-statistic. If there is a unit-root, then it is concluded that the series is generated from a nonstationary process (*e.g.,* random walk with trend process). Otherwise, if there is no unit-root but there is a trend, then it means that the series is generated from a trend-stationary process.

---

[2] Although the rule of thumb is to use 0.05 or 0.01 levels of significance in hypothesis testing, in this study 0.10 is used for unit-root testing.

Second, in the case that there is no trend given that there is a unit-root, a second regression is estimated but excluding the trend term. Subsequently, a unit-root test is performed. If there is a unit-root, then a joint test for the significance of the constant term and a unit-root is carried out. If the constant term is not significant, then the third step is executed. Otherwise, if the constant term is significant, then a unit-root test is implemented using the z-statistic. If there is a unit-root, then the test suggests that the series is generated from a nonstationary process (*e.g.,* random walk with drift process). Otherwise, if there is no unit-root, then it is concluded that the series is generated from a stationary process.

Third, if both trend and constant terms are not significant given that there is a unit-root, then a third regression is estimated without these two terms. Subsequently, a unit-root test is carried out. If there is a unit-root, then it implies that the series is generated from a nonstationary process (*e.g.,* random walk). Otherwise, if there is no unit-root, then it is concluded that the series are generated from a stationary process.

The estimation of the above three equations is crucial in testing the significance of a unit-root. According to Enders (1995) the presence of additional regressors that are not part of the DGP may significantly reduce the power of the tests, and thus lead to accepting the null hypothesis that there is a unit-root when it is false (type II error).

1.4.1.1 Objective Two

Trends and systematic effects are usually present in times-series. To eliminate these effects the actual yield data are generally filtered. However, in the agricultural economics literature, a clear empirical guidance on how the yield series must be transformed prior to density estimation is not provided. This thesis proposes to examine the possible impact of using linear-detrending and first-differencing filtering techniques on crop yield density estimation. Towards this end, actual crop yields are appropriately and inappropriately filtered according to the identified DGP; such that two additional series are produced by applying linear-detrending and first differences to the actual corn yield data. The next step is to estimate two crop yield density functions via the nonparametric kernel method using the filtered yield data (residuals). Percentile estimates are used to represent probability estimates. The final step is the computation of percent errors which are the presumed difference between the incorrect and correct probabilities divided by the correct probabilities. A magnitude of zero percent error means no effect, while a greater or lower than zero magnitude is indicative of an overestimation or underestimation, respectively.

1.5 Thesis Outline

This thesis is divided into five chapters. Chapter 1 consists of an introduction, which provides background information on the contemporary issues of probability density function estimation, a problem definition that leads to the present study, a justification of why this research is needed, the objectives that provide specific synopsis of the research purpose, and the subsequent thesis outline. Chapter 2 provides literature review on data filtering techniques and crop yield density estimation. Chapter 3 presents the methodology needed to conduct the study. Chapter 4 includes the results obtained from the analysis. Chapter 5 provides the summary, implications and conclusions.

2. LITERATURE REVIEW

This chapter summarizes recent contributions to crop yield density estimation. This literature review is organized in four sections. The first gives a preamble to basic concepts of trends and their modeling techniques. The second provides examples of agricultural risk modeling works that apply filtering techniques. The third presents the two main approaches (parametric and non-parametric) as well as the different distributions that can be used in density estimation. The fourth and last section summarizes the findings.

2.1 Time-Series Background

A time-series is stationary if its mean and variance do not change over time (Hamilton, 1994). However, most time-series used in economics are rarely generated from a stationary process, as they often present either a deterministic or stochastic trend. Transformations of the raw data are usually carried to meet the stationary condition. Detrending and first differencing are transformations (filters) commonly used in the field of agricultural economics.

Linear trends have been used in previous works (*e.g.,* Gallagher, 1986; 1987; and Taylor, 1990). The approach consists in estimating a simple regression model with a trend as an independent variable; the residuals from the model are used to fit a density function. This approach is suitable whenever crop yield series present a constant rate of growth over time, which appears to be the case in some production areas.

Stochastic trends are defined as processes in which the mean and variance are changing over time or as shocks that have slowly decaying effects (Hamilton, 1994). The interaction of a farm's natural amenities (*e.g.,* soil fertility and weather) with technology (*e.g.,* new machinery and new inputs) results in crop yields that increase over time although at different growth rates across farms. Systematic effects also occur due to droughts, floods, and other influences

(Goodwin and Ker, 1998). The recent tendency to account for trends is to take the first differences of the crop yield observations or the use of ARIMA filters (*e.g.,* Goodwin and Ker, 1998).

The most appropriate technique to filter a trend-stationary (deterministic trend) series is detrending, and for a stochastic trend process (random walk) it is first differencing (Hamilton, 1994). Hamilton (1994) stated that "a final difference between trend-stationary and stochastic trend process (unit-root) that deserves comment is the transformation of the data needed to generate a stationary time-series" (Hamilton, 1994, p. 444). Enders (1995) also insisted on this important fact suggesting that the stochastic properties of the series are determinant in choosing the appropriate transformation to achieve stationarity. Nelson and Kang (1981) illustrated the consequences of inappropriate transformations and in particular the spurious periodicity implied by detrending a random walk series via a Monte Carlo experiment. Hamilton (1994) and Enders (1995) recommended that deciding on whether a process is better represented by a deterministic or stochastic trend (unit-root process) can be determined through unit-root testing.

According to the theoretical literature (Hamilton, 1994 and Enders, 1995) on time-series, if a yield series has a unit-root, then applying first differences produces a stationary series. However, problems may arise when a trend-stationary process (equation 1.1) is first-differenced. Hamilton (1994) showed that if the true model for yields is trend-stationary:

$$(1.1)\ \ y_t = \alpha t + y_0 + \varepsilon_t,$$

then first differencing introduces a unit-root (the $-\ \varepsilon_{t-1}$ term) into the moving average component, that is,

$$(1.2)\ \Delta y_t = \alpha t + y_0 + \varepsilon_t\ -\ \varepsilon_{t-1}.$$

Another empirically relevant point is that, considering a random walk with drift model, if the series is detrended, then the time-dependence of the mean is successfully eliminated, but not the variance, thus failing to render the series stationary (Hamilton, 1994).

It is well-known that non-stationarity can cause spurious regressions (Enders, 1995; Hamilton, 1994; and Hill, Griffiths and Lim, 2008), meaning that the regression appears to be significant when in fact it is not. In other words, variables may be mistakenly found to be highly correlated. Beyond this crucial issue, the cost of not assessing the appropriateness of data transformations may be high, particularly in crop yield density estimation. It has become apparent, then, that probability density functions and the resulting probability estimates may be unreliable if an inappropriate transformation is used.

2.2 Modeling Trends in Crop Yields

Early works on crop yield density estimation used actual yield series, also referred to as levels, to estimate density functions in crop insurance ratemaking (Botts and Boles, 1958). When carrying out analyses using actual yield time-series data, the stationary condition is assumed to be met. However, Foote and Bean (1951) suspected a trend in U.S. corn yields (1866-1949) which may have started around the 1940s and that the fluctuations around the suspected trend were not random. Two types of series were tested for randomness. The first is a series of observed per-acre yields and the second, a series that was built by introducing “several repetitions of a trend, a 13 year pattern, and an extraneous element added at 10 year interval” (Foote and Bean, 1951, p. 24). However, the test suggested that both series were random when in fact the second was nonstationary due to the introduced effects. The conclusion was that the procedure available at that time (*e.g.,* a test based on deviations from moving averages, and deviations from a trend based on the theory of runs) appeared to be ineffective at the moment in distinguishing between random and nonrandom processes.

Day (1965) also attempted to demonstrate empirically the presence of trends, using experimental data on corn, oats, and cotton from 1921 to 1957 from the Mississippi River Delta. However, after carrying out the Wald-Wolfowitz test (based on the theory of runs), he concluded that crop yields were “more or less random.” He was cautious on his conclusion, perhaps because of the demonstrated low reliability of the test as suggested by Foote and Bean. Additionally, Day emphasized that it is necessary to attain a better understanding of the stochastic properties of field crop yields in order to convert uncertainty into risk efficiently. This statement is still true today and has implications for important risk analyses in agriculture.

Fitting linear trends to yield data became common practice later on. Luttrell and Gilbert (1976) analyzed the randomness of U.S. wheat, corn, rye, barley and oats yields via the Wallis-Moore test for cyclical fluctuations using data for the 1866-1970 period which was divided into two sub-periods: 1866-1932 and 1933-1970. In the first sub-period, Luttrell and Gilbert found a significant trend by carrying out a regression of the logarithm of yields on simple time and time squared trends. However, evidence of “bunchiness” (stochastic trends) was not revealed. In the second period, a non-significant trend was obtained. This result was probably a consequence of the choice of structural change which was based solely on a visual inspection of the series over time. In addition, the occurrence of a severe drought corresponding to the Dust Bowl period (1930s-1940s) and the high rate of nitrogen adoption period (about 1950s) may have produced more than one structural change in the series, making linear trend modeling even more difficult.

Learning from the Lutrell and Gilbert’s work, Gallagher (1986) proposed to separate U.S. corn yield series based on technology adoption. The periods were divided into low technology adoption (1933-1955) and rapid technology adoption (1955-1960). Subsequently, he estimated and forecasted corn yields under the capacity concept employing linear and logarithmic trends to account for low and rapid technology adoption, respectively. The corn yield estimates were tested for randomness, as a reliability examination, using the nonparametric Wald-Wolfowitz and the Wallis-Moore tests. The tests suggested the non-presence of phases or cycles. However, this outcome may have been different if a more powerful test such as the augmented Dickey Fuller (1979) would have been implemented. Also, Gallagher (1987) studied U.S. soybean yields for the 1941-1984 period, but at this time using only a single linear trend, and arrived at the same conclusions as those for corn.

Houck and Gallagher (1976) studied the responsiveness of U.S. corn yields (1951-1971) to input and output prices. They regressed corn yields as a dependent variable on lagged USDA weighted average fertilizer price indexes, nationwide harvested acres, weather during the growing season, a dummy variable representing U.S. corn acreage restrictions and corn blight during 1970 and linear and logarithmic trends accounting for the effects of technical improvements. Although they found both trend types to be significant, the logarithmic was more intuitive to model yield improvements at a decreasing rate. Additionally, Houck and Gallagher stated that the reaction of corn prices is underestimated when fertilizer prices are held constant. This finding suggests that an analysis of corn production variation should include the effect of changes in corn and fertilizer prices.

Menz and Pardey (1983) tackled technology dividing it into non-nitrogen and nitrogen components for U.S. corn yields during the 1954-1980 period. Their approach differed from the one taken by Houck and Gallagher in replacing the lagged relative fertilizer price indexes by the natural logarithm of nitrogen application rates. Menz and Pardey found yearly yield increments not attributed to nitrogen to be around one bushel per acre and that same rhythm was expected to continue in the near future. Also in opposition to the results obtained by Houck and Gallagher, Menz and Pardey found a diminishing relationship between corn yields and nitrogen prices. They concluded, that "no yield plateau has yet been reached" (Menz and Pardey, 1983, p. 561).

One of the first empirical and Monte Carlo studies illustrating the relevance of non-stationarity in crop yield data is that of Zapata and Rambaldi (1989). They found that arbitrarily chosen transformations such as detrending, first differencing, and log changes to commodity prices and yields can produce series with very different properties than those of the underlying stochastic process. They also suggested that the ambiguity on deciding which data transformation technique to use can be solved by identifying the DGP of the series through unit-root testing. Surprisingly, however, the empirical literature to date on crop yield trend and risk modeling has remained silent on this issue until recently (*e.g.,* Sherrick *et al.,* 2004).

Nelson and Preckel (1989) continued addressing the issue of corn yield responsiveness to nitrogen prices using Iowa experimental data for the 1961-1970 period. Although they departed from previous work (*e.g.,* Houck and Gallagher, 1976 and Menz and Pardey, 1983) that employed detrending, they did not consider the effect of a trend in estimating crop yield density functions. In a later publication Nelson recognized the importance of detrending (Nelson, 1990). Nevertheless, he did not consider any type of detrending, because of working with short-term time-series data (county yield data for the 1964-1969 period).

Taylor (1990) provided alternative procedures under small sample size data for estimating crop yield density functions. The empirical analysis used corn, soybean and wheat yields from Macoupin County, Illinois, for the 1945-1987 period, reporting the use of linear-detrending. Turvey and Zhao (1999) ranked the performance of several distributions utilizing detrended spring grain, wheat, corn, soybean and white bean yield data from 609 farms in 10 counties. Turvey and Zhao used detrended data because their source "the Ontario crop insurance commission" had already adjusted it (Turvey and Zhao, 1999, p. 8). However, in the above two works, as in many others, the failure to consider stochastic trends may have led to serious estimation bias. Furthermore, by implementing linear-detrending, they may have likely introduced shocks into the yield series (Hamilton, 1994).

Just and Weninger (1999) utilized alfalfa (1960-1993), corn (1941-1994), sorghum (1980-1994), soybean (1969-1994), and wheat (1926-1994) county level yields in estimating trends. They stated that the usual assumption of linear-detrending was perhaps the cause of having more evidence against normality. They also suggested the use of polynomial trends (up to the fifth order) to better depict the different changes in the rate of yield improvements. Conversely to Thompson (1969), Just and Weninger justified the choice of the polynomial order by carrying out F-tests and using the Akaike Information Criterion (AIC). Gommes (2006) reported the use of a "curvilinear trend," but without specifying the polynomial order employed when forecasting Zimbabwe's corn yields. In the above works, polynomials are well suited in the case of trend-stationary yield series, but these approaches may not approximate stochastic trends which are a dominant property of crop yield series.

Throughout the literature, it is noted that the use of deterministic and polynomial trends has been so vastly employed that their inclusion in multiple agricultural economics sub-disciplines such as risk analysis and production economics has become somewhat mechanical. An exemption was the formal consideration of stochastic trends in Kaylen and Koroma (1991). They modeled corn yield and weather data for the 1895-1988 through a "univariate state space time-series" model, a method that nests deterministic and stochastic trends possibilities. Moss and Shonkwiller (1993) followed a similar approach, but using farm level yields with a shorter time-series 1930-1990. However, empirical evidence as to whether yield models exhibit stochastic or deterministic trends was not provided.

Goodwin and Ker (1998) added new dimensions to the evolution of this literature. They introduced a univariate filtering model, an ARIMA (0,1,2) to best represent crop yield series. The use of this universal filter is intuitive in the sense that the model accommodates a unit-root (*i.e.* the number "1" in ARIMA (0,1,2)) and an infinite autoregressive structure through the 2-MA term. Thus, mathematically, the approximation should work well for a large number of "nonstationary" processes. One limitation to this work is that the DGP has been generalized to all crop county yield series combinations. This choice was justified on practical grounds, but the DGP may not be a generalization that applies to all county yields and, as a consequence, the ARIMA filter may have been inappropriate in some instances. Examples of cases when it may be inappropriate are with deterministic trending yields.

Atwood, Shaik and Watts (2003) presented a Monte Carlo study to assess the effect of detrending short-term panel-data on normality tests. Trended observations were simulated to accommodate no slope and steeper slope models. The fitted trends were (1) no trend adjustment, (2) Just and Weninger's approach of estimating individual farm trends and (3) error component implicit detrending (ECID). The results suggested that normality tests tend to be biased in a type II direction when estimating individual trends for short term panel data. Consequently, Atwood, Shaik and Watts strongly agreed with Just and Weninger in that incorrect detrending and not accounting for heteroskedasticity was perhaps a cause of bias in normality test in previous works (Atwood, Shaik and Watts, 2003). However, Atwood, Shaik and Watts cautioned researchers when accounting for trends and heteroskedaticity in short term panel data, because not accounting for these effects may produce better performance in normality tests. They demonstrated that an implication of individual detrending and assuming normality is a reduction of the costs of crop insurance *premia*. While this Monte Carlo experiment revealed several important empirical points, there are uncertainties about the effects of filtering yield series not generated from deterministic processes that have not been resolved. This gap can be filled with the simulation of stochastic trends and ARIMA models.

Sherrick *et al.* (2004), as in Zapata and Rambaldi (1989), assessed DGP identification by carrying out unit-root tests (Phillips-Perron test). The empirical analysis utilized corn and soybeans data for the 1972-1999 period from the University of Illinois endowment farms in 12 counties with at least 20 observations. They found all 26 corn yield series to be trend stationary and only one from the 25 soybean yield series with a unit-root, a clear counter-example to the validity of ARIMA filters. Consequently, linear and polynomial trends were fitted and their selection was based on the significance of F-tests as done in Just and Weninger's works. Interestingly enough, their results are inconsistent with what was expected according to conclusions in previous works (a dominance of stochastic trend processes over deterministic processes, *e.g.*, Goodwin and Ker, 1998). Thus, it seems relevant to continue exploring this subject by studying methods that would allow a conclusion with validity on the presence and type of trends in crop yield series. It is also important to mention that Foote and Bean (1951) wanted a test capable of detecting the somewhat suspected non-randomness in crop yields. It was not until the work of Zapata and Rambaldi (1989) and Sherrick *et al.* (2004) that unit-root tests were employed in the literature of agriculture economics, even though the test have been available since mid-1970s.

2.3 Crop Yield Density Estimation

2.3.1 Parametric Density Estimation

Early applications of crop yield density estimation in risk assessments were based on the estimation of a normal distribution (*e.g.,* Botts and Boles, 1958). Although the non-normality (skewness) of crop yields was questioned in Foote and Bean (1951), it was not until Day (1965) that an empirical alternative was reached through a beta density function. Day's forecast was for positively skewed yields. He hypothesized that weather constraints such as drought, too much water, and warmer or colder conditions than usual would reduce yield by a greater magnitude than the increase generated due to occurrence of excellent weather throughout the growing season. Based on that, he stated that "less than average yields are more likely than greater than average yields" (Day, 1965, p. 714). But Day's assumption of positively skewed yield densities did not hold in all cases. He found negatively skewed beta density functions in the case of oats. Another important contribution of this work was to show the usefulness of crop yield densities conditioned on inputs. It was interesting to learn how the responsiveness of yields to different nitrogen levels was revealed by the change in the shape of the densities. Day concluded that crop yield series are non-normal and that the skewness and kurtosis degree is linked to crop type and nutrients. These conclusions were the foundation to a subject matter that has been of continued interest ever since.

In agreement with the theory of skewed crop yields, but with a different perspective, Gallagher (1986) pointed out that skewed densities from U.S. corn average yields are more likely since biological capacity cannot be surpassed and can approach zero under blight or early frost because producers are not independent in space, they share the same area and similar weather conditions. Gallagher confirmed this assumption by finding negatively skewed gamma distributions in corn yields for the 1933-1981 period. He emphasized that the "actuarial soundness" of crop insurance programs based on the normal distribution can be affected because of the possible underestimation of the high chances of obtaining an extremely low yield or the likelihood of over passing average yields (Gallagher, 1986). The same conclusions were reached for average U.S. soybean yields for the 1941-1984 period (Gallagher, 1987).

Negative skewness in crop yield densities caught the attention of Nelson and Preckel (1989). They evaluated the parametric beta distribution to depict the corn yield distribution response to input applications. According to Nelson and Preckel, the beta distribution provides better estimates than non-parametric density estimation methods when correct assumptions about the distributions are made. They also stated that the beta distribution has some advantages for agricultural analyses because it allows for a variety of shapes, which can be either negatively or positively skewed. The beta distribution also permits the incorporation of several explanatory variables, but leaves unchanged the number of parameters to be estimated. However, the cost of those mentioned advantages is diminished flexibility. They used farm level yield data from five Iowa counties, but only for those years when corn was grown. Other characteristics such as soils (lodging, percentage clay in soil, soil type, and soil slope), applied inputs (nitrogen, phosphates, and potassium), insecticides, and rotations were also recorded. These were duly evaluated, as they can influence skewness and variance in the conditional beta distribution. Maximum likelihood was employed in two stages. They estimated unconditional and conditional distribution parameters. In the latter, α and β functional forms were selected *a priori*. In conclusion, Nelson and Preckel found that generally inputs either augmented or reduced the mean, skewness and variance. In opposition, they did not consider that yields could be correlated among farms, heteroskedasticity, and that significance skewness cannot be sustained if normality is not rejected (Just and Weninger, 1999).

Nelson (1990) was the first to address the issue of the impact of the distribution choice on probability estimates. In particular, he demonstrated how *a priori* wrong assumption of a normally distributed crop yields can overestimate an insurance premium. He estimated and compared crop insurance *premia* at three coverage levels (50%, 65% and 75%) based on normal and gamma distributions. He found significantly larger *premia* to be charged to farmers based in the former case.

Taylor (1990) brought for the first time the use of the multivariate non-normal probability density function that can work under a small sample size (*e.g.,* farm data). He found both univariate and multivariate densities for corn, soybeans and wheat to be negatively skewed. This work was the motivation for other multivariate studies (*e.g.,* Ramirez, 1997) who also reported negative skewness. Moss and Shonkwiler (1993) also found significant evidence of non-normally and negatively skewed corn yield distributions through the estimation of an inverse hyperbolic sine transformation (IHST). In spite of having an almost closed debate regarding negative skewness, Just and Weninger (1999) pointed out that the normality rejection was attributable to three common mistakes (1) inappropriate use of average time-series data to infer farm yield distributions, (2) linear detrending (3) and beta distributions specification that assumes negative skewness when yields are normal. Additionally, Just and Weninger affirmed that under their proposed approach which is the use of farm or experimental data, polynomial trends and normal distribution, normality results are seldom rejected. At this point, again, the small consensus reached at least in terms of skewed crop yields is challenged. Nevertheless, "as recognized by the authors of previous studies, nonrejection does not prove yield normality, because the magnitudes of the type-two errors in their normality tests are unknown" (Ramirez, Misra and Field, 2003, p. 119).

The results of Sherrick *et al.* (2004) suggested that the beta distribution surpassed the fitting performance of weibull, logistic, normal and lognormal in density estimation of corn and soybeans. Estimated expected insurance premium prices from the five distributions show significant ranges. Sherrick *et al.,* therefore, suggested that further research should include the analysis of nonparametric methods.

2.3.2 Nonparametric Density Estimation Techniques

Goodwin and Ker (1998) employed a semi-parametric approach with a nonparametric kernel smoother. The authors were motivated by the flexibility advantage of the kernel method methodology which does not require previous knowledge of the functional form, supports different density shapes (*e.g.,* skewness and kurtosis) and let local idiosyncrasies be more realistically represented compared to the parametric approach. Pooled cross-sectional yield data at the county level were used. This data arrangement was chosen in order to add information from neighboring counties (weighted for the county being modeled) and avoid low performance by the nonparametric kernel method when using small sample sizes. Goodwin and Ker found negative skewness, as had many previous authors. In addition, as a result of nonparametric estimation, new elements are revealed that would otherwise be omitted. Slight evidence of bimodality was found in seven of the eight counties, so yields close to the maximum attainable occur frequently while the opposite happen constantly too. In conclusion, they demonstrated that by using nonparametric techniques, the results obtained are quite different compared to those employing parametric methods; and may therefore lead to improved probability estimates and more accurate Group Risk Plan (GRP) premium rates.

Turvey and Zhao (1999) ranked the performance of Normal, Gamma and Beta (parametric distributions) and kernel density method (non-parametric) using five different crop farm-level yields. The kernel method produced *premia* with minimum error more than 50 percent of the time. Turvey and Zhao concluded that, in the context of crop insurance premium estimation, the non-parametric approach is the most efficient. Also, Norwood, Roberts and Lusk (2004) focused on the predictive ability of an entire distribution and ranked Gamma, Beta, Stochihs, Multihs, semi-parametric (kernel), and Normal distributions through the Out-of-Sample-Log-Likelihood function (OSLLF). Norwood Robert and Lusk assured that OSLLF was the criterion to use because the purpose was to obtain information on both interpolative and extrapolative forecasts. Their results bring new elements that may close the debate around the distribution choice. Norwood Robert and Lusk found that the semi-parametric (kernel) approach of Goodwin and Ker (1998) has the greatest performance in forecasting county average yields. This conclusion constitutes the motivation for using non-parametric methods in the present research, in the context of unit-root and deterministic trend processes. Light is added to uncover the effect of filtering procedures on density functions via percent errors.

2.4 Summary

The review of literature suggests several empirically relevant points (questions). First, do unit-root tests on time-series of corn and soybean yields support deterministic trends or stochastic trends? Second, do arbitrarily chosen time-series filtering techniques have an impact on crop yield density function used in risk analysis? How much discrepancy exists in the calculated probabilities of two alternative density functions for the same yield data? To what extent are the findings sensitive to model structure, nonstationary properties of yields and sample size? The next chapters help reach answers to most of these questions. The main proposition is that if an assessment of data filtering is conducted through unit-root tests, then the chances of proper filtering yield data are enhanced, leading to more reliable yield density estimation.

3. METHODOLOGY

The general procedure to estimate a density function for crop yields typically requires the application of a data filter prior to density estimation. The previous chapter identified the types of filters used in previous work, namely, detrending, first-differences or ARIMA modeling. Since the true data generating process (DGP) is never known in practice, it is possible that arbitrary filtering of crop yields may result in a misidentified density function.

Table 3.1 presents some possible scenarios (first column) when DGP is unknown (second column) and alternative filters are used (third column). For example, in fitting a crop yield probability density function, with no prior knowledge about the DGP, choosing first-differences is the correct choice of filter in case the case of a random walk (fourth column), but if instead an incorrect filter is employed, as shown in the second scenario of Table 3.1, the probability estimates may differ. In order to resolve this issue, a methodology that is capable of identifying the proper filter is needed.

The methodology proposed in this thesis for choosing the "right-filter" prior to yield density estimation uses existing statistical tests for nonstationary time-series. More specifically, "unit-root" tests are proposed because such tests can accommodate yield data properties consistent with the scenarios in Table 3.1 (*e.g.,* linear and stochastic trends). The approach consists of the following steps. First, identify the DGP that is more consistent with yield data using unit-root tests; second, determine the appropriate filtering method and estimate yield probabilities; and third, calculate percent errors in yield probabilities by comparing the best filter to an alternative filter.

Table 3.1. Data Transformation Scenarios.

Scenario	Data Generating Process	Transformation	Type of Transformation
I	Random Walk	Differencing	Right
II	Random Walk	Detrending	Wrong
III	Trend Stationary	Detrending	Right
IV	Trend Stationary	Differencing	Wrong
V	Stationary	None	Right
VI	Stationary	Differencing	Wrong
VII	Stationary	Detrending	Wrong

Note: the random walk, random walk with drift and random walk with trend are reported as "Random Walk" since the correct filter (differencing) to achieve stationarity is the same for all cases.

3.1 Identification of the Data Generating Process

This empirical analysis uses per county corn and soybean yield data from Arkansas and Louisiana with a yield history range of 30 to 49 years (1960-2008). A basic step in the identification of the DGP is to plot the series against time; to the trained eye this provides a visual idea of the underlying process. The visual inspection of the plots, however, is not reliable enough, so formal testing is conducted. In this thesis unit-root testing is carried on historical corn and soybean county yields to identify whether a yield series is generated from a stationary or nonstationary process. More specifically, the augmented Dickey-Fuller (ADF) test is employed through a procedure suggested by Enders (1995) when the DGP is not known. The ADF testing procedure is based in the estimation of three equations (steps) that are used to test the presence of (1) a unit-root and a trend, (2) a unit-root and a drift and (3) a simple unit-root. Figure 3.3 presents these three steps in yellow, orange and green areas which illustrate the flow of the entire procedure along with information on the equations, hypotheses ($e.g., Ho: \gamma = 0$) and tests statistics ($e.g.,$ φ_3) needed for the hypothesis testing. All unit root tests are conducted at the 0.10 level of significant because of the low power of unit-root tests documented in the literature.

The first step is to estimate the most general equation of the ADF test (equation 3.1) by ordinary least squares. This is a regression of first-differences in yields (Δy_t) on a constant term

(α_0), a linear trend ($a_2 t$), a lagged yield (γy_{t-1}) and a lagged differences ($\sum_{s-1}^{m} \beta_i \Delta y_{t-i}$). The first-differences and the lagged difference terms result from subtracting a lagged yield term from both sides of the equation and account for the presence of autoregressive effects; the presence of a constant term and a linear trend is tested because they can be present in stochastic processes; and as many as necessary lagged differences terms are estimated to capture "the full dynamic nature of the process" and to fix the possible autocorrelation of the residuals (Hill, Griffiths and Lim, 2008, p. 337). The optimal number of lags is chosen by ranking models based on the lowest estimate for the Akaike's information criterion (Akaike, 1974). The equation including all these above described terms is given by

$$(3.1)\ \Delta y_t = \alpha_0 + \alpha_2 t + \gamma y_{t-1} + \sum_{s-1}^{m} \beta_i \Delta y_{t-i} + \varepsilon_t,$$

where ε_t denotes a random error term. Subsequently, the null hypothesis that a yield series is generated from a unit-root or nonstationary process ($Ho\colon \gamma = 0$) is tested using τ_τ statistic at the 0.10 significance level. The alternative hypothesis is that yield series is not generated from a unit-root process or yield series is stationary ($H_1\colon \gamma \neq 0$). If a unit-root is not detected, then it is concluded that the series is stationary. Otherwise, if there is a unit-root, then a joint test is carried to test for the significance of the trend and the presence of a unit-root ($Ho\colon a_2 = \gamma = 0$) using the φ_3 statistic. If the trend is not significant, then the second step is conducted. Otherwise, if the trend is significant, then again a test for the presence of a unit-root using z-statistic is conducted. If there is a unit-root, then it is concluded that the yield series is generated from a random walk with trend process. Otherwise, if there is no unit-root, it is concluded that the series is generated from a trend-stationary process.

In the second step, if no trend is found but there is a unit-root, a regression (equation 3.2) is estimated without the trend term:

$$(3.2)\ \Delta y_t = \alpha_0 + \gamma y_{t-1} + \textstyle\sum_{s-1}^{m} \beta_i \Delta y_{t-i} + \varepsilon_t.$$

Subsequently, a unit-root test is carried using a τ_π statistic. If there is no unit-root, then it is concluded that the yield series is stationary. Otherwise, if there is a unit-root, then a joint test for the significance of the constant term and a unit-root is carried ($Ho\colon a_0 = \gamma = 0$) using φ_1 statistic. If the constant term is not significant, then the third step is carried. Otherwise, if the constant term is significant, then again a test for the presence of a unit-root using z-statistic is conducted. If there is a unit-root, then it is concluded that the series is generated from a random walk with drift process. Otherwise, it is concluded that the yield series is generated from a stationary or white noise process.

Third, if no constant and no trend are found but there is a unit-root, then another regression (equation 3.3) is estimated without these terms as given by

$$(3.3)\ \Delta y_t = \gamma y_{t-1} + \textstyle\sum_{s-1}^{m} \beta_i \Delta y_{t-i} + \varepsilon_t.$$

Subsequently, a unit-root test is carried using a τ statistic. If there is no unit-root, then it is concluded that the series is generated from a stationary or white noise process. Otherwise, if there is a unit-root ($H_1\colon \gamma \neq 0$), then the conclusion is that the series comes from a nonstationary process (*e.g.,* random walk).

The estimation of the above three equations is crucial in testing the significance of a unit-root. According to Enders (1995) the presence of regressors that are not part of the DGP may significantly reduce the power of the tests, and thus lead to accept the null hypothesis when in reality is false (type II error). Hence, it is important to estimate and re-test for the presence of a unit-root with a regression that does not include non-significant terms as done in the previous steps.

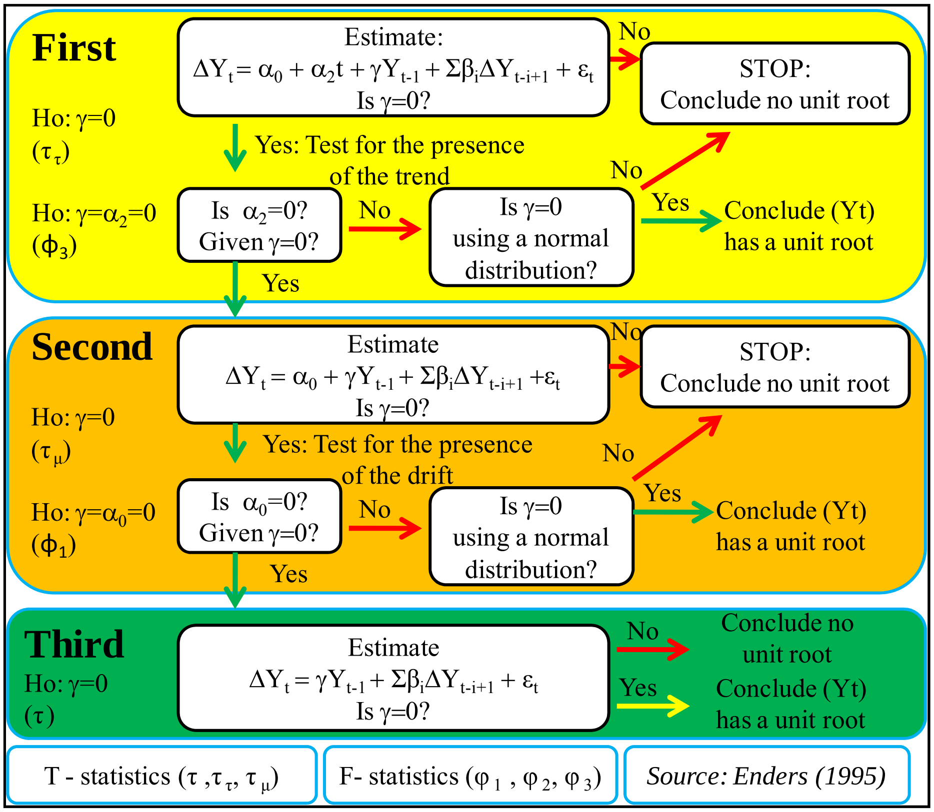

Figure 3.1. Augmented Dickey-Fuller Test. Source: Enders (1995).

3.2 Alternative Filtering Techniques

In order to assess the effect of alternative filtering on empirical density functions, two sets of alternatively transformed data are produced for each crop yield series. Hamilton (1994) stated that the appropriateness of a filter to achieve stationarity depends on whether series comes from a trend-stationary process or has a unit-root. More specifically, Hamilton recommended that if a nonstationary series (unit-root) behaves like a random walk model which is given by

$$(3.4)\ y_t = y_{t-1} + \varepsilon_t,$$

then a stationary series is generated by taking the first-differences of the series

$$(3.5)\ \Delta y_t = \varepsilon_t;$$

otherwise, if the yield series is trend-stationary given by

$$(3.6)\ y_t = a_1 + a_2 t + \varepsilon_t,$$

then a stationary series is obtained by regressing the yield series on a simple function of time

$$(3.7)\ y_t - (a_1 + a_2 t) = \varepsilon_t;$$

and in the case that the yield series is generated from a stationary process (white noise), no data transformation is needed. In this thesis, these recommendations are used in deciding the correct data filter to apply.

Once the series has been appropriately transformed, an alternative filter is applied following the proposed scenarios shown in Table 3.1. These filters are chosen based on Hamilton (1994) who cautions towards inappropriate data transformations which are: if a trend-stationary series is first-differenced, then a unit-root is introduced into the process and if a random walk process is detrended, then the time dependency of the trend is solved but not the growing variance. These new transformed series are used in the estimation of a density function in the following step.

3.3 Non-parametric Density Estimation

Density functions are used in crop yield risk analyses and can be generated from time-series data. As pointed out earlier the use of alternative approaches to transform data may have an effect on density functions. Therefore a density is estimated using appropriately and inappropriately filtered data through the non-parametric kernel method (using the Gaussian density as the kernel). The kernel estimator of a probability density function of a sequence of identically and independently distributed yield observations is represented by

$$(3.8)\quad \hat{f}(y) = \frac{1}{\sum_{i=1}^{n} W_i} \sum_{i=1}^{n} W_i\, \varphi h\, (y - Y_i),$$

where $\hat{f}(y)$ is the kernel density approximation of a probability density function of the yield series, Y is the crop yield series, W is a symmetric weighting variable and h is the smoothing parameter which determines the amount of weight of the neighboring observations in density estimation. A large h parameter creates a smooth density, whereas a small generates a bumpy density. Selecting the smoothing parameter is crucial in nonparametric kernel density estimation. For a detailed review of the "rule-of-thumb" method used in this thesis for estimating the smoothing parameter please refer to Silverman (1986).

One advantage of using non-parametric over parametric methods is that the identified density function is driven by the data itself, and this flexibility eliminates the *a priori* restriction that the data should conform to some known parametric functional form. Furthermore, the majority of previous studies found negatively skewed and kurtotic crop yield density functions; thus the flexibility of the non-parametric method would represent such shapes.

3.4 Percent Errors

If a probability density function is affected by filtering, then its probability estimates are suspect. One attractive approach to measure the effects of filtering is to compute the percent error produced between alternative probabilities. [3] In this study percentile estimates (25^{th}, 50^{th}, and 75^{th}) are used to represent probability estimates. The percent error (PE) is given by

$$(3.9)\ PE = ((PI - PC)/PC) * 100,$$

where *PI* is the percentile of a density estimated from incorrectly filtered series, *PC* is the percentile of a density estimated from correctly filtered data. The absolute value of a percent error is used to represent the magnitude of the impact of using inappropriate data transformations, while the sign of a percent error can be interpreted as an underestimation or overestimation of the probability estimates of yields. To illustrate, Figure 3.2 shows two

[3] The two-sample Kolmogorov-Smirnov-test (K-S) is a plausible method to determine whether or not two differently filtered series come from the same distribution. This test requires series to be independent. Portmanteau test for independence suggested that the two alternatively filtered yield series are dependent (the tabulated p-values are available from the author upon request). Thus, in this particular case the K-S test cannot be implemented. In view of this limitation and that to our best knowledge there is no test capable of revealing the differences between two dependent time-series samples, a percent error computation is proposed to assess the impact of alternative filters in crop yield density functions.

probability density functions (PDF) that were estimated using appropriately and inappropriately filtered data. Besides the difference in the density shapes, it is clear that at the 25th percentile in the "appropriately-filtered-series PDF" predicts a lower yield (green area of Figure 3.2) than the "inappropriately-filtered-series PDF" (blue area plus green area of Figure 3.2) and that difference is called a "percent error."

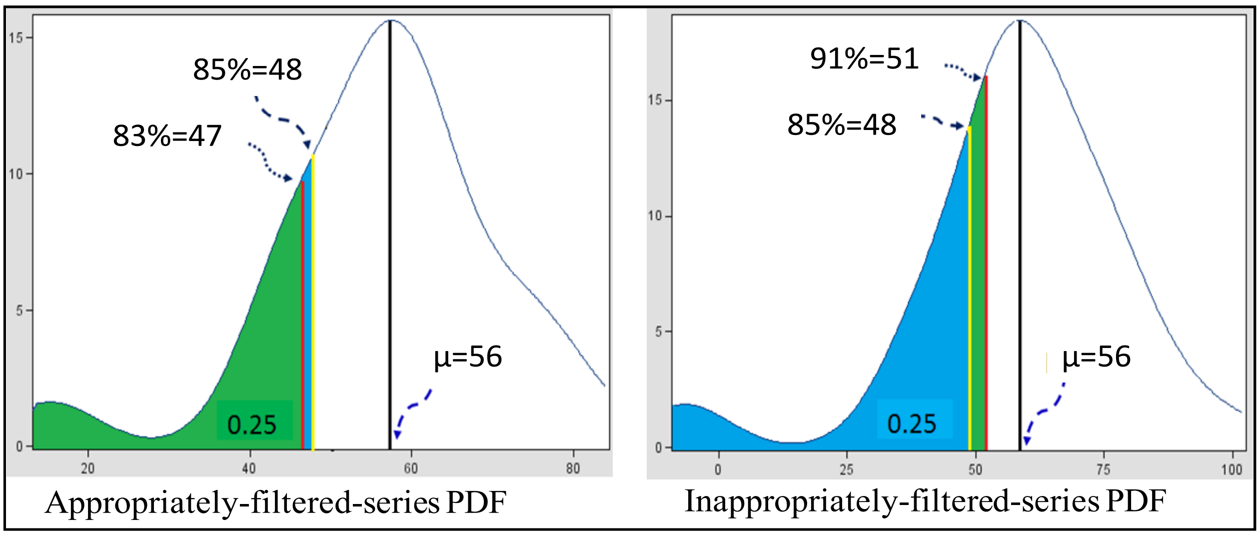

Figure 3.2. Percent Error from Alternative PDFs.

4. RESULTS

This chapter describes the results obtained from applying unit-root tests, using the augmented Dickey-Fuller approach, to corn and soybean yields in Arkansas and Louisiana for the 1960-2008 period. Unit-root test are used to identify whether crop yields can be best represented as purely random around a constant mean and variance (stationary) or by some other model (*e.g.*, a simple linear trend or a random walk process). The main proposition of this study was that such tests can remove the uncertainty often associated with deciding how to filter crop yields prior to density estimation. In this chapter, the first section presents a descriptive analysis of the corn and soybean yield data in Arkansas and Louisiana. In the second section, the augmented Dickey-Fuller unit-root test is applied by estimating equations 3.1-3.3 in chapter 3. This section includes the identification of the data generating process for all county yields. The third and last section of this chapter addresses the nonparametric density estimation of probability estimates. This section reports an assessment of the appropriateness of filters in the form of percent errors between alternative probabilities that were estimated from alternative filters.

4.1 The Data

Historical corn and soybean yields (bu/acre) for Arkansas and Louisiana were obtained from the National Agricultural Statistics Service for the 1960-2008 period. The data are aggregated county level yields for irrigated and non-irrigated crops for counties with a production history of at least 30 years and no more than five years of missing data. These data screening resulted in a total of 31 corn and soybeans producing counties in Arkansas and 25 corn and 34 soybeans producing parishes in Louisiana.

4.1.1 Descriptive Analysis of Corn Yields in Arkansas

Descriptive statistics for corn yields in Arkansas are shown in Table 4.1. The state level average yield was 85.6 bu/acre from 1960-2008 with the highest yield of 160 bu/acre in 2007 and the lowest of 26 bu/acre in 1964, respectively. The state level yields for corn had a standard deviation of 43.2 bu/acre. County level yields for corn were highest at 191 bu/acre in Lonoke County in 2007 and lowest at 17 bu/acre in Conway in 1964. The highest standard deviation for county yields was 53.37 bu/acre in Prairie, while the lowest was 31.9 bu/acre in Miller.

Table 4.1. Descriptive Statistics of Corn Yields in Arkansas.

County	N. Observations	Mean	Maximum	Minimum	Standard Dev.
Arkansas State Total	48	85.56	169	26	43.22
Arkansas	36	84.18	180	22	50.04
Ashley	36	69.53	165	28	41.17
Chicot	45	78.87	166	22	44.50
Clay	48	100.27	176	27	45.42
Conway	46	74.12	155	17	44.35
Craighead	48	89.73	184	24	45.09
Crawford	30	57.19	135	20	33.77
Crittenden	41	85.00	165	30	43.27
Cross	35	67.12	136	28	35.50
Desha	44	89.14	187	32	47.05
Drew	32	64.88	173	21	45.25
Greene	48	87.34	172	24	41.62
Independence	38	78.10	160	25	44.51
Jackson	48	91.00	159	24	43.74
Jefferson	44	87.59	177	31	47.69
Lafayette	44	79.04	173	24	40.76
Lee	40	86.25	176	25	50.50
Little River	37	67.36	146	24	35.48
Logan	30	70.29	150	22	46.38
Lonoke	39	86.02	191	22	53.13
Miller	43	66.50	120	20	31.90
Mississippi	48	93.70	173	28	44.08
Monroe	39	86.11	162	26	49.46
Phillips	40	79.89	176	26	48.77
Poinsett	44	85.73	180	24	41.70
Prairie	43	90.16	175	25	53.37
Randolph	46	88.17	173	25	47.58
Saint Francis	36	78.84	173	28	48.15
White	41	67.88	139	23	36.96
Woodruff	46	87.95	153	28	41.43
Yell	40	66.79	136	18	36.53

4.1.2 Descriptive Analysis of Corn Yields in Louisiana

Descriptive statistics for Louisiana corn yields are shown in Table 4.2. The state level average yield was 84.29 bu/acre from 1960-2008 with the highest of 163 bu/acre in 2007 and the lowest of 27 bu/acre in 1960. The state level yields for corn had a standard deviation of 39.28 bu/acre. County level yields were highest at 184.8 bu/acre in Natchitoches Parish in 2007 and lowest at 17 bu/acre in Washington Parish in 1960. The highest standard deviation for county yields was 49.62 bu/acre in Franklin Parish, while the lowest was 14.2 bu/acre in Allen Parish.

Table 4.2. Descriptive Statistics of Corn Yields in Louisiana.

County	N. Observations	Mean	Maximum	Minimum	Standard Dev.
Louisiana State Total	49	84.29	163	27	39.28
Acadia	37	61.64	130	22	32.17
Allen	35	44.22	70	24	14.20
Avoyelles	49	87.26	161	25	39.78
Beauregard	41	56.46	100	28	23.65
Bossier	41	60.70	163	23	30.78
Caddo	49	74.73	169	21	37.90
Catahoula	47	79.18	160	25	35.43
Concordia	49	80.58	158	31	32.92
East Carroll	48	87.84	172	36	40.31
East Feliciana	41	62.60	119	21	22.86
Evangeline	37	67.65	163	22	36.45
Franklin	49	93.65	180	22	49.62
Madison	49	84.64	158	35	34.45
Morehouse	48	86.10	179	25	48.86
Natchitoches	48	78.06	185	23	43.46
Ouachita	37	69.98	144	27	35.28
Pointe Coupee	49	94.31	177	30	44.19
Rapides	49	98.48	158	34	37.79
Red River	44	71.67	160	21	37.50
Richland	49	76.02	164	21	42.80
Saint Landry	49	87.24	166	27	36.89
Tangipahoa	40	64.64	123	18	25.37
Tensas	49	85.24	157	35	36.34
Washington	42	62.58	116	17	25.72
West Carroll	49	72.45	173	19	43.51

4.1.3 Descriptive Analysis of Soybean Yields in Arkansas

Descriptive statistics for soybean yields in Arkansas are presented in Table 4.3. The state level average yields was 25.7 bu/acre from 1960-2008 with the highest yield of 39 bu/acre in 2003 and the lowest yield of 15 bu/acre in 1980. The state level yields for soybeans had a standard deviation of 5.75 bu/acre. County level yields for soybeans were highest at 46 bu/acre in Arkansas County in 2004 and lowest at 8 bu/acre in Dallas County in 1980. The highest standard deviation was 7.96 bu/acre in Randolph, while the lowest was 3.60 bu/acre in Nevada.

Table 4.3. Descriptive Statistics of Soybean Yields in Arkansas.

County	N. Observations	Mean	Maximum	Minimum	Standard Dev.
Arkansas State Total	47	25.66	39	15	5.75
Arkansas	49	31.79	46	19	6.80
Ashley	49	22.98	37	11	6.41
Clark	39	22.94	33	12	4.26
Clay	49	24.80	41	13	7.56
Conway	45	23.11	34	11	5.27
Craighead	49	25.97	41	13	6.98
Cross	49	27.80	42	15	6.64
Dallas	32	18.69	30	8	4.50
Desha	49	25.66	43	14	7.02
Faulkner	46	22.33	37	11	5.78
Hempstead	37	22.91	32	11	3.86
Independence	49	22.98	37	10	4.96
Jackson	49	22.25	34	12	5.64
Jefferson	49	25.77	38	13	5.92
Lawrence	45	23.87	37	13	6.58
Lee	49	25.25	40	17	5.72
Logan	46	24.03	40	11	5.84
Lonoke	49	27.69	42	14	6.36
Miller	39	21.53	38	14	4.47
Mississippi	49	27.74	43	17	6.63
Monroe	49	25.04	39	16	5.96
Nevada	36	20.80	29	11	3.60
Perry	42	22.44	37	12	5.73
Phillips	48	25.40	41	14	6.17
Prairie	49	28.79	43	16	6.56
Pulaski	49	24.19	38	11	5.98
Randolph	49	24.87	41	13	7.96
Saline	34	19.00	33	12	4.38
Sebastian	43	23.48	35	13	5.37
White	49	22.07	33	11	5.71
Yell	46	21.97	35	11	5.65

4.1.4 Descriptive Analysis of Soybean Yields in Louisiana

Table 4.4 shows the descriptive statistics for Louisiana soybean yields. The state average yield was 26.3 bu/acre from 1960-2008 with the highest yield of 43 bu/acre in 2007, the lowest yield of 20 bu/acre in 1964 and a standard deviation of 4.9 bu/acre. Parish level yields were high at 51.9 bu/acre in Iberville in 2007 and lowest at 7.6 bu/acre in West Carroll in 1980. The highest standard deviation was 7.9 bu/acre in Iberville, while the lowest was 4.4 bu/acre in Evangeline.

Table 4.4. Descriptive Statistics of Soybean Yields in Louisiana.

County	N. Observations	Mean	Maximum	Minimum	Standard Dev.
Louisiana State Total	49	26.31	43.00	20.0	4.90
Acadia	49	26.23	40.10	13.0	5.05
Allen	45	22.38	30.80	11.9	4.68
Avoyelles	49	27.40	45.00	18.4	5.01
Beauregard	44	24.09	34.60	14.9	5.56
Bossier	45	24.03	40.40	10.9	5.55
Caddo	49	23.66	40.00	15.0	5.54
Calcasieu	41	23.91	33.00	13.8	4.31
Caldwell	46	22.42	35.00	12.5	5.42
Cameron	37	23.68	32.50	14.3	3.98
Catahoula	49	24.26	45.80	16.7	5.32
Concordia	49	26.21	51.00	14.5	6.19
East Carroll	49	28.54	48.40	14.5	7.87
Evangeline	47	24.69	34.30	15.7	4.38
Franklin	49	22.01	39.20	11.2	6.51
Grant	35	26.37	37.20	10.6	5.99
Iberia	37	29.53	47.30	14.0	7.53
Iberville	43	33.48	51.90	20.0	7.87
Jefferson Davis	47	25.54	37.30	12.0	4.76
Lafayette	43	27.12	34.50	16.8	4.75
Madison	49	26.60	41.90	14.2	5.87
Morehouse	49	23.00	41.10	10.5	6.30
Natchitoches	40	25.36	50.00	11.0	7.75
Ouachita	48	22.99	38.60	13.7	5.72
Pointe Coupee	44	33.51	51.00	19.0	7.08
Rapides	49	30.30	46.50	18.5	5.76
Red River	44	24.28	35.60	13.3	5.40
Richland	48	21.02	46.80	9.5	7.62
Saint Landry	49	27.09	40.00	18.2	4.46
Saint Martin	45	26.42	39.40	14.0	4.97
Tensas	49	27.49	50.80	15.8	6.73
Vermilion	44	25.58	39.40	11.30	4.56
West Baton Rouge	39	32.45	46.50	19.10	5.38
West Carroll	48	22.90	38.20	7.60	7.12
West Feliciana	38	25.00	34.40	15.20	4.60

4.1.5 Arkansas and Louisiana Corn and Soybean yields General Overview

As expected, due to the technological improvements, most low yields are observed in the 1960s, whereas high yields occurred during the first decade of the 2000s. Notice that the maximum and minimum corn and soybean yields at the state and county level occurred nearly during the first and last decade of the studied period, respectively. Higher yields are concentrated around the Mississippi River Delta, an area known for its nutrient rich soils. This common yield performance at the state level might be attributable to the fact that the most productive areas are located in the Mississippi River Delta. The descriptive analysis reveals that corn yields in Arkansas have higher averages at the state and county level and higher state level standard deviation than in Louisiana. Contrarily, soybean yields in Louisiana present higher averages at the state and county level and a slightly lower state level standard deviation than in Arkansas.

Time plots of corn and soybean yields in Lonoke and Grant counties in Arkansas and Louisiana, respectively are shown in Figure 4.1. As a general observation, the majority of counties had crop yields that are characterized by a seemingly deterministic upward trend (left-hand side of Figure 4.1). However, there are instances when yields were more random around a constant mean (right-hand side of Figure 4.1). The information that can be extracted from a plot may help not only to understand the general behavior, but also to have a better intuition about the data generating process (DGP). To illustrate, the left-hand side corn yield series in Figure 4.1 presents a clear upward trend. One may interpret the series as a trend-stationary process with a deterministic-trend characteristic of a constant rate of technological progress in yields. Others may categorize the same yield series as a stochastic-trend because of the evident changes in its slope and variance throughout the 1960-2008 period. Both types of processes can be broadly described as nonstationary. Considering the right-hand side in Figure 4.1, the soybean yield series seems to be randomly fluctuating around a constant rather than presenting a clear upward trend. This leads one to think that the series may have been generated from a stationary process. However, visual inspection is not reliable enough, and thus formal testing should be included.

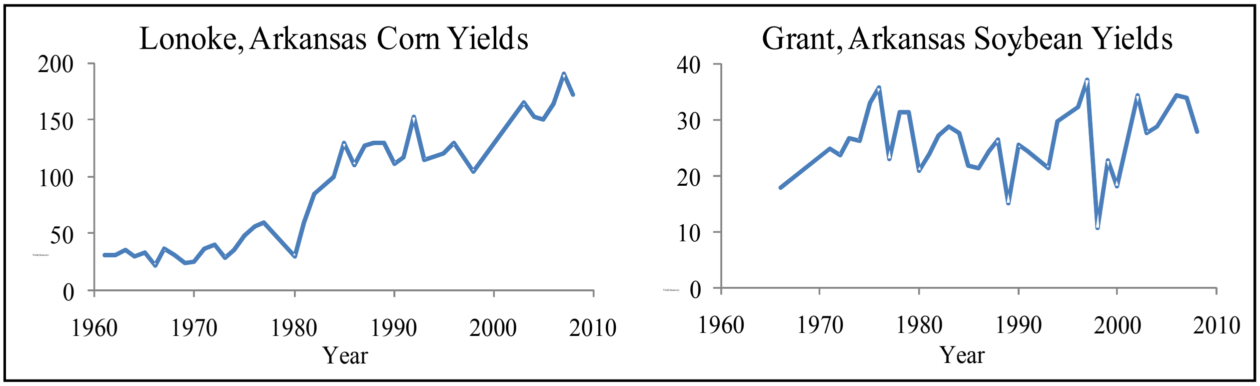

Figure 4.1. County Crop Yields.

4.2 Filtering Yield Data

The usual first step in the estimation of crop yield density functions is to obtain a random series with no systematic components. Because most crop yields do not meet such a requirement, it is usual practice to filter the data prior to density estimation. The augmented Dickey-Fuller (ADF) test for unit-root is used to identify the DGP that best fits corn and soybeans county yields.

Table 4.5 presents selected results of the unit-root tests. [4] Even though not all the test statistics are reported to save space, the purpose of Table 4.5 is to illustrate how the root-methodology should be implemented in identifying the DGP for corn and soybean yields presented in Table 4.6. For the first case, Allen Parish, Louisiana, corn yields are tested for the presence of a unit-root based using the estimated regressions given by equations 3.1-3.3 for the ADF statistic of chapter 3 (first to fifth row in Table 4.5). The first five rows of table 4.5 report the results. The least restrictive regression (equation 3.1, chapter 3) includes a constant and a

[4] Complete tabulated unit-root results (121 tables) are available from the author upon request.

trend terms and is estimated in order to test for the presence of a unit-root. The resulting test statistic for the null hypothesis (Ho: γ=0) is -2.33 with a critical value (τ_τ) of -3.13 (first row, Table 4.5).[5] At the 10% level of significance the null hypothesis is not rejected (rejection values are those smaller than -3.13) meaning that the series has a unit root. [6] Therefore, as explained in chapter 3, a joint test for the significance of the trend given that there is a unit root is conducted. The resulting test statistic for the null hypothesis (Ho: γ=α_2=0) is 2.73 with a critical value (φ_3) of 5.34 (second row in Table 4.5). The implied hypothesis is that there is no linear trend (Ho: α_2=0) but a unit-root maybe present, thus, this hypothesis is not rejected, implying that corn yields do not follow a deterministic trend but there may be a unit-root present. A conclusion on the presence of a simple unit root cannot be made at this stage because additional regressors that are not part of the DGP may decrease the power of the test (Enders, 1995). Thus, equation 3.2 is estimated in order to test the presence of a unit-root and a constant term (third row, Table 4.5). The resulting test statistic for the null hypothesis of a unit-root (Ho: γ=0) is -1.20 and the critical value (τ_μ) is -2.57 which suggest that there is a unit-root. Subsequently, a joint test for the presence of a constant term given the presence of a unit-root is performed (fourth row, Table 4.5). The test statistic for the null hypothesis (Ho: γ=α_0=0) is 1.05 and the critical value (φ_1) is -3.78. These values suggest rejecting the null hypothesis, so the constant term is not significant. Still, it is not possible to conclude about the presence of a unit-root, as previously suggested the estimation of a regression including additional regressors (in this case a constant term) that are not part of the DGP may bias the test leading one to accept the null hypothesis when in it is not true (Enders, 1995). Therefore, equation 3.3 is estimated (fifth row, Table 4.5). The resulting test statistic for the null hypothesis (Ho: γ=0) is 0.44 with the critical value (τ) of -1.62.

---

[5] The same critical value is used for the test based in equation 3.1.
[6] The same significance level is used for all unit-root tests.

Consequently, the null hypothesis of a unit-root is not rejected, leading to a conclusion that the series has a unit-root. One example of a series that presents a unit-root is a random walk process; and after looking at a plot of Allen's actual corn yields against time, it is noticeable that these yields follow a very similar behavior as the one of a random walk (*e.g.,* left-hand side of Figure 4.1). Therefore, after conducting this systematic sequence of five hypothesis tests, the main conclusion is that Allen corn yields are characterized by a random walk process as reported in Table 4.6. Note that it is tempting to think, by visual inspection, that corn yields on the left hand side of Figure 4.1 follow a simple linear trend. In fact, if a linear trend regression is estimated for such data, it may be significant. This may have misled initial efforts in crop yield density estimation to filter the data with a simple linear trend.

Second, in the case of Bossier Parish, Louisiana corn yields (sixth row, Table 4.5) equation 3.1 of chapter 3 was estimated in order to conduct the unit-root tests. The resulting test statistic for the null hypothesis (Ho:γ=0) that there is a unit-root is -3.68 with a critical value (τ_τ) of -3.13. Therefore, the null hypothesis should be rejected and there is no unit-root. A better estimate can be obtained with a joint test of a unit-root and no linear trend as done in case 1 (row 7, Table 4.5). The resulting test statistic for the null hypothesis (Ho: γ=α_2=0) is 6.84 with a critical value (φ_3) of 5.34. The conclusion is that there is no unit root and that corn yields in Bossier Parish, Louisiana, are trend deterministic.

In the third example (eighth row, Table 4.5), a unit-root test is carried out for soybean yields in Grant Parish, Louisiana. Equation 3.1 of chapter 3 was estimated and the resulting test statistic for the null hypothesis (Ho: γ=0) was -3.56 with a critical value (τ_τ) of -3.13. This result suggests that there is no unit-root and there is no need to proceed further. Therefore, the conclusion was that the series is stationary (as presented in Table 4.6). Observe that the right-hand side plot in Figure 4.1 shows no upward trending behavior. In fact, a flat line located somewhere between 20 and 30 bu/acre would represent the data well; this would be the yield mean which Table 4.4 shows is 26.37 bu/acre.

Table 4.5. Test Statistics and Critical Values of the ADF Unit-Root Test.

State	County	Crop	Equation	Test Statistic	Asy. Critical Value (10%)	Hypothesis	Statistic
Louisiana	Allen	Corn	3.1	-2.33	-3.13	$\gamma=0$	τ_τ
Louisiana	Allen	Corn	3.1	2.73	5.34	$\gamma=\alpha_2=0$	φ_3
Louisiana	Allen	Corn	3.2	-1.20	-2.57	$\gamma=0$	τ_μ
Louisiana	Allen	Corn	3.2	1.05	3.78	$\gamma=\alpha_0=0$	φ_1
Louisiana	Allen	Corn	3.3	0.44	-1.62	$\gamma=0$	τ
Louisiana	Bossier	Corn	3.1	-3.68	-3.13	$\gamma=0$	τ_τ
Louisiana	Bossier	Corn	3.1	6.84	5.34	$\gamma=\alpha_2=0$	φ_3
Louisiana	Grant	Soybeans	3.1	-3.56	-3.13	$\gamma=0$	τ_τ

Note: $\gamma=0$ is the null hypothesis of a unit-root and uses τ_τ statistic, $\gamma=\alpha_2=0$ is the joint hypothesis of a unit-root and a trend and uses φ_3 statistic and $\gamma=\alpha_0=0$ is the null hypothesis of a unit root and a constant term and uses φ_1 statistic.

Analyses similar to those for cases 1-3 were conducted for each corn and soybean producing county in Arkansas and Louisiana included in the study (121 tables not reported here). The results of the identified processes are summarized in Table 4.6. The column labeled DGP (data generating process) can also be conveniently labeled "Filter." For aggregated state level yields in Arkansas, for example, yields behave as a random walk (RW). Therefore, the appropriate filter for these yields is to use first-differences. Notice that the process that best explains yields varies from county to county. This means that the assumption that different types of processes such as random walk, trend-stationary and stationary can be found in crop yields.

Table 4.6. Crop Yields Series Data Generating Process as Suggested by Unit-Root Testing.

Arkansas				**Louisiana**			
Corn		Soybeans		Corn		Soybeans	
County	DGP	County	DGP	County	DGP	County	DGP
Arkansas State Total	RW	Arkansas State Total	RW	Louisiana State Total	RW	Louisiana State Total	RW
Arkansas	RW	Arkansas	RW	Acadia	RW	Acadia	RW
Ashley	TS	Ashley	TS	Allen	RW	Allen	ST
Chicot	RW	Clark	RW	Avoyelles	RW	Avoyelles	RW

Table 4.6. Continued.

Clay	RW	Clay	RW	Beauregard	RW	Beauregard	TS
Conway	TS	Conway	TS	Bossier	TS	Bossier	RW
Craighead	RW	Craighead	RW	Caddo	RW	Caddo	TS
Crawford	RW	Cross	TS	Catahoula	RW	Calcasieu	ST
Crittenden	RW	Dallas	RW	Concordia	RW	Caldwell	RW
Cross	RW	Desha	RW	East Carroll	RW	Cameron	ST
Desha	RW	Faulkner	RW	East Feliciana	TS	Catahoula	RW
Drew	RW	Hempstead	RW	Evangeline	TS	Concordia	RW
Greene	RW	Independence	TS	Franklin	RW	East Carroll	RW
Independence	RW	Jackson	TS	Madison	RW	Evangeline	RW
Jackson	TS	Jefferson	RW	Morehouse	TS	Franklin	RW
Jefferson	RW	Lawrence	TS	Natchitoches	TS	Grant	ST
Lafayette	RW	Lee	RW	Ouachita	RW	Iberia	RW
Lee	RW	Logan	RW	Pointe Coupee	RW	Iberville	TS
Little River	RW	Lonoke	TS	Rapides	RW	Jefferson Davis	RW
Logan	RW	Miller	RW	Red River	TS	Lafayette	ST
Lonoke	RW	Mississippi	RW	Richland	RW	Madison	RW
Miller	RW	Monroe	RW	Saint Landry	TS	Morehouse	RW
Mississippi	RW	Nevada	RW	Tangipahoa	RW	Natchitoches	RW
Monroe	RW	Perry	RW	Tensas	TS	Ouachita	RW
Phillips	RW	Phillips	RW	Washington	RW	Pointe Coupee	RW
Poinsett	RW	Prairie	RW	West Carroll	TS	Rapides	RW
Prairie	TS	Pulaski	RW			Red River	RW
Randolph	TS	Randolph	RW			Richland	RW
Saint Francis	RW	Saline	RW			Saint Landry	RW
White	RW	Sebastian	RW			Saint Martin	TS
Woodruff	RW	White	TS			Tensas	RW
Yell	RW	Yell	RW			Vermilion	RW
						West Baton Rouge	RW
						West Carroll	RW
						West Feliciana	RW

Labels: ST stands for stationary, TS for trend-stationary, RW for random walk and DGP for data generating process.

A summary of the results in Table 4.6 using percentages is shown in Figure 4.2. There were a total of 121 yield series for both states. In the case of corn, unit-root testing suggests that yield data are generated from unit-root processes in 84% and 64% of the counties studied for Arkansas and Louisiana, respectively, while the remaining 16% and 36% of the cases were generated from a trend-stationary process. In the case of soybeans, unit-root tests identified a yield stochastic trend process in 73% of the counties studied, while in the remaining 29% and 12% a trend-stationary process was suitable for Arkansas and Louisiana, respectively. Note that in 15% of the cases, Louisiana soybean yield series could be represented by a stationary process.

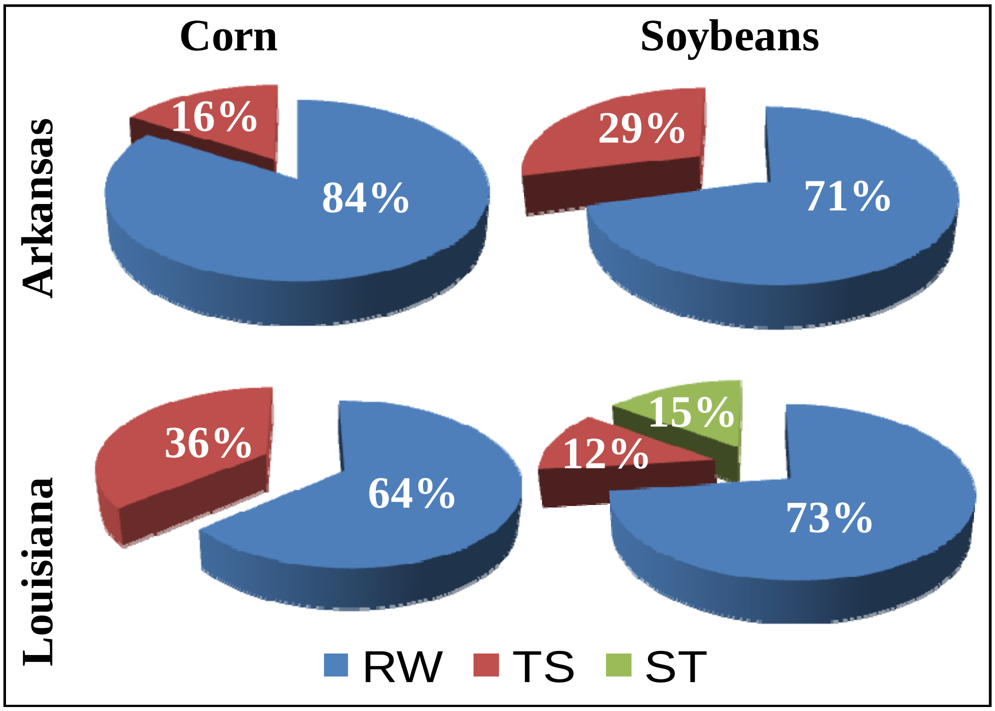

Figure 4.2. Summary of Processes for Crop Yield Series.

Figure 4.3 presents a geographical distribution of the identified DGP for corn and soybean yields. The maps on the upper-level of the Figure 4.3 correspond to the state of Arkansas, while the ones on the lower-level are for the state of Louisiana. The left-hand side maps correspond to corn producing counties, while the ones on the right-hand side to soybean producing counties. The diagonal gridded red pattern within the counties describes yields that follow a random walk, the green dark color denotes trend-stationary yields and the vertical/horizontal gridded light blue pattern represents stationary crop yields. Notice that random walk processes tend to form clusters around the areas where the state standard deviations for yields are high, for example, the area around the Mississippi River Delta. Corn yields series that are trend-stationary, on the contrary, are less likely to cluster together. Soybeans yields that are trend-stationary tend to group at the interior of the two states (more at the Southwest in the case of Louisiana). The counties presenting stationary crop yield series represent only 4% of the total studied counties (121) from the two states combined, and this process is found only for soybeans producing counties in Louisiana. This last resulting process (stationary) is not surprising due to the little or no yield improvement (at the county level) observed over the entire studied period at those locations. Figure 4.3 clearly shows that there are crop yields from the same areas that can be generated from different DGP; therefore, the use of a single filter to all the series prior to density estimation is clearly inappropriate for these data.

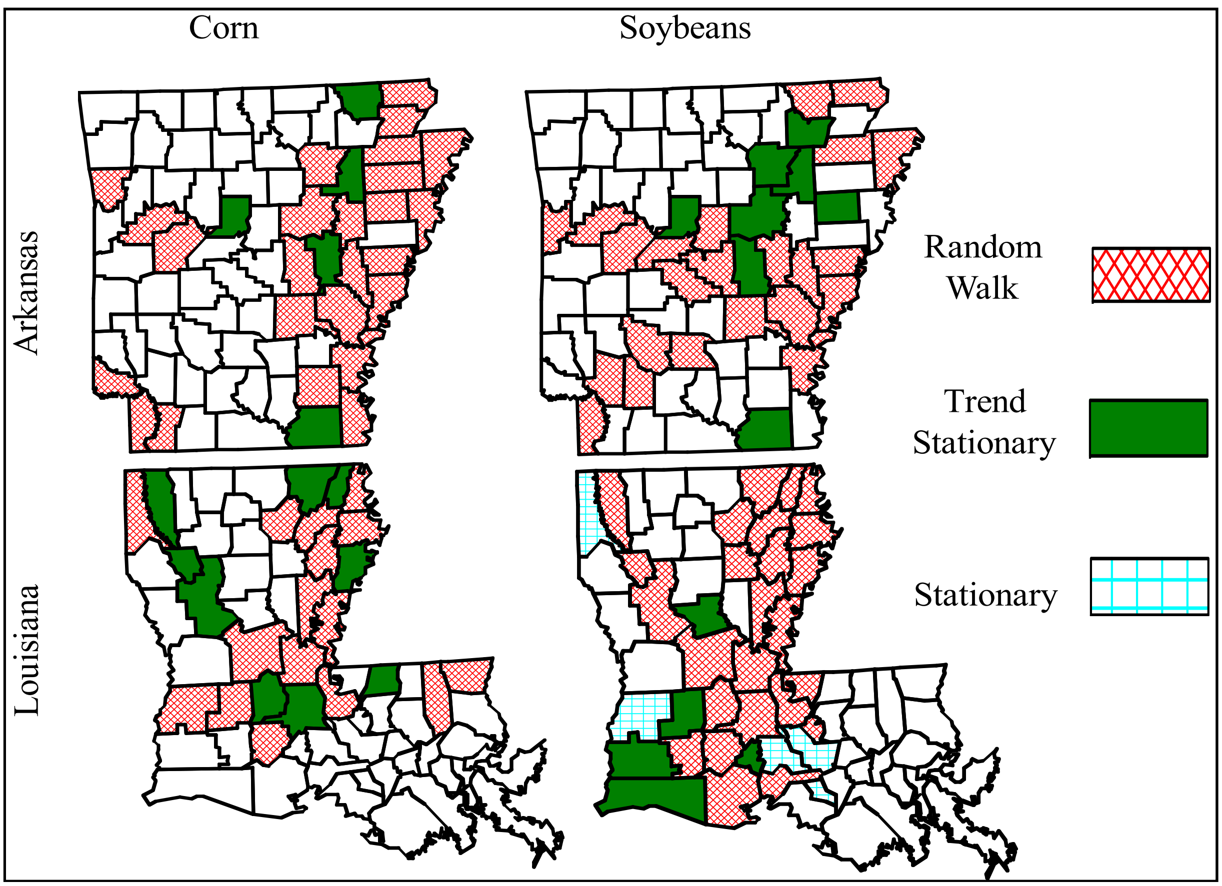

Figure 4.3. Appropriate Filters for Crop Yields in Arkansas and Louisiana.

4.3 Non-parametric Kernel Density Estimation Results

One of the approaches most widely used in measuring crop yield risk is the estimation of probability density functions. The probability associated to the realization of an expected yield is used to determine the level of crop yield risk. A yield "probability estimate" is usually represented as the area under the curve from the left tail to point of the expected yield in the probability density function. In this analysis, density percentiles are used to represent "probability estimates." The resulting percentile estimates are shown in Tables 4.7, 4.8, 4.9 and 4.10 for Arkansas corn and Louisiana corn and in Arkansas soybeans and Louisiana soybeans, respectively. These Tables are divided into two sections labeled "Random Walk" and "Trend Stationary" depending on the result of the unit-root tests for each county yield series.

The "Random Walk" section has a first column labeled "County" that contains the list of counties in which crop yield series were identified as a random walk process. The second column labeled "Differenced RW" shows the percentiles from a distribution estimated from appropriately transformed yield data, in this case, first differencing a random walk series. The third column labeled "Detrended RW" contains the percentiles from a distribution estimated from inappropriately transformed yield data, in this case, detrending a random walk series. The fourth column labeled "Percent Error" presents the percentile differences between distributions based on appropriately and inappropriately filtered data, computation which is given by equation 3.9. For example, in Table 4.7, the percentiles from a distribution based on appropriately transformed corn yield series for Arkansas County, Arkansas are -4.20 (25^{th}), 0.30 (50^{th}) and 12 (75^{th}); the percentiles from inappropriately transformed data are -14.52 (25^{th}), 3.02 (50^{th}) and 15.47 (75^{th}); and the percent error are 245.78% (25^{th}), 908.12% (50^{th}) and 28.88% (75^{th}).

The “Trend Stationary” section of the Table 4.7, for example, has a first column labeled “County” that contains the list of counties for which crop yield series are trend-stationary. The second column labeled “Detrended TS” shows the percentiles from a distribution estimated from appropriately transformed yield data, in this case, detrending trend-stationary series. The third column labeled “Differenced TS” contains the percentiles from a distribution estimated from inappropriately transformed yield data, in this case, differencing trend-stationary series. The fourth column labeled “Percent Error” presents the percentile differences between distributions based on appropriately and inappropriately filtered data, which computation is the “Differenced TS” percentiles minus the “Detrended TS” percentiles dived by the “Detrended TS” percentiles. For example, the percentiles in Table 4.7 from a distribution based on appropriately transformed corn yield series for Ashley County, Arkansas, are -14 (25th), 1.99 (50th) and 10.91 (75th); the percentiles from inappropriately transformed data are -8 (25th), 3 (50th) and 10 (75th); and the percent errors are -42.86% (25th), 50.83% (50th) and -8.34% (75th). Additionally, percentiles and percent errors for aggregated state level yields are also calculated.

4.3.1 Arkansas Corn Yields

Percentiles and percent errors from densities estimated from appropriately and inappropriately filtered corn yields for Arkansas are presented in Table 4.7. The effect of inappropriate filtering yield data (detrending a random walk) prior to estimating a corn yield density function produces on average a density function with the following percent errors: 33.22% (25th), -132.78% (50th), and -10.75% (75th) which are located in the first row and fourth column in Table 4.7. At the county level, the effect of inappropriate filtering is even larger, for example, the percent error when detrending a random walk corn yields in Arkansas County is 908.12% (50th), while the lowest in Clay County is -0.44% (25th). In this particular case, percent errors reveal that overestimation and underestimation would happen in equal proportions when detrending a random walk yield-series (the wrong filter). On the contrary, first-differencing a trend-stationary yield series (the incorrect filter) produces on average an overestimation in 60% of the cases. These are important errors that need to be considered, especially when designing crop risk management programs based on the distribution of crop yields as the judgments on the most probable yields can be biased. The probability of a loss is a very important policy parameter in insurance ratemaking. Thus, in order to better represent the impact of alternative filtering methods in such programs, an application of the assessment of a probability loss under alternative probabilities estimates provided later in the thesis in chapter 5.

Table 4.7. Arkansas Corn Yield Kernel Probability Estimates in Percentile Values.

County	**Random Walk**								
	Differenced RW			**Detrended RW**			**Percent Error (%)**		
	25th	**50th**	**75th**	**25th**	**50th**	**75th**	**25th**	**50th**	**75th**
Arkansas State Total	-6.00	2.00	11.00	-7.99	-0.66	9.82	33.22	-132.78	-10.75
Arkansas	-4.20	0.30	12.00	-14.52	3.02	15.47	245.78	908.12	28.88
Chicot	-8.00	1.50	13.00	-9.66	-0.49	10.36	20.76	-132.78	-20.30
Clay	-8.30	3.00	14.00	-8.26	-0.48	7.07	-0.44	-115.96	-49.52
Craighead	-6.10	1.00	13.00	-10.39	0.55	11.19	70.33	-44.56	-13.95
Crawford	-5.00	0.40	11.00	-14.16	0.60	13.36	183.19	50.54	21.42
Crittenden	-7.50	2.00	10.50	-11.72	3.70	12.61	56.21	85.07	20.11
Cross	-3.00	4.00	8.90	-11.37	0.33	11.76	279.03	-91.76	32.10
Desha	-9.00	2.30	16.00	-13.65	2.63	10.55	51.64	14.16	-34.03
Drew	-8.00	4.00	18.00	-14.41	-0.31	13.69	80.09	-107.68	-23.95
Greene	-5.00	2.70	13.00	-8.20	1.16	8.18	63.93	-57.08	-37.06
Independence	-8.00	1.00	12.60	-15.97	1.04	17.51	99.63	3.54	38.96
Jefferson	-8.00	2.00	14.50	-11.41	0.66	11.61	42.62	-66.86	-19.90
Lafayette	-5.00	4.00	13.00	-11.77	-1.42	7.70	135.35	-135.43	-40.81
Lee	-6.10	1.00	12.60	-14.85	1.10	13.19	143.38	9.98	4.66
Little River	-7.50	1.95	13.00	-9.79	-0.47	13.25	30.57	-124.15	1.90
Logan	-6.00	3.00	11.00	-20.82	2.55	19.47	246.99	-15.04	77.03
Lonoke	-7.00	3.50	13.00	-13.00	-2.96	17.95	85.77	-184.44	38.05
Miller	-8.00	2.00	12.00	-10.45	-2.42	10.78	30.66	-221.12	-10.17
Mississippi	-10.00	3.00	13.00	-8.41	-0.15	13.41	-15.88	-104.92	3.12
Monroe	-7.00	2.65	13.00	-15.64	-2.06	12.00	123.43	-177.58	-7.69
Phillips	-6.00	3.30	11.00	-14.13	-1.06	19.08	135.49	-132.09	73.47
Poinsett	-6.00	2.00	12.00	-6.58	1.36	6.70	9.59	-32.21	-44.18
Saint Francis	-8.00	3.00	13.00	-11.73	-0.18	14.99	46.66	-106.10	15.29
White	-7.00	4.50	10.75	-9.45	-1.21	10.93	35.03	-126.93	1.69
Woodruff	-7.00	0.00	10.00	-8.45	-0.36	8.82	20.66	36.02	-11.80

Table 4.7. Continued.

Yell	-6.00	4.00	12.50	-8.60	0.31	7.37	43.37	-92.25	-41.01
	Trend Stationary								
	Detrended TS			**Differenced TS**			**Percent Error (%)**		
Ashley	-14.00	1.99	10.91	-8.00	3.00	10.00	-42.86	50.83	-8.34
Conway	-7.12	-0.04	11.17	-10.00	0.00	13.00	40.42	-100.00	16.41
Jackson	-8.29	0.62	11.47	-10.00	0.00	13.00	20.60	-100.00	13.33
Prairie	-12.73	-0.16	13.22	-5.00	1.00	12.00	-60.72	-741.79	-9.24
Randolph	-12.24	1.66	12.94	-9.00	1.40	13.00	-26.49	-15.86	0.43

Note: Differenced RW is obtained by taking the first differences of an I(1) yield series (the correct filter); Detrended RW are the residuals from removing the a simple linear trend from a RW yield series (the wrong filter); Detrended TS are the residuals from removing a simple linear trend from a TS yield series (the correct filter); and Differenced TS is obtained by taking the first differences of a TS yield series (the wrong filter).

4.3.2 Louisiana Corn Yields

Percentiles and percent errors from densities estimated from appropriately and inappropriately filtered corn yields for Louisiana are presented in Table 4.8. The effect of detrending a random walk yield series (the wrong filter) produces on average density function with the following percent errors 6.31% (25^{th}), -144.8% (50^{th}), and -17% (75^{th}) which are located in the fourth column and first row in Table 4.8. At the parish level, the highest impact as measured by the percent error was -1,153.65% (50^{th}) in East Feliciana Parish, while the smallest for Natchitoches Parish 1.14% (25^{th}). Percent errors show patterns that vary with the type of transformation carried out. For instance, underestimation would be induced by detrending a random walk series (the incorrect filter) in 67% of the cases, while an overestimation is produced when differencing a trend-stationary series (the incorrect filter) in 56% of the cases.

Table 4.8. Louisiana Corn Yield Kernel Probability Estimates in Percentile Values.

County	**Random Walk**								
	Differenced RW			**Detrended RW**			**Percent Error (%)**		
	25^{th}	**50^{th}**	**75^{th}**	**25^{th}**	**50^{th}**	**75^{th}**	**25^{th}**	**50^{th}**	**75^{th}**
Louisiana State Total	-8.00	2.00	10.00	-8.50	-0.90	8.30	6.31	-144.80	-17.00
Acadia	-7.20	3.85	10.00	-11.64	0.13	13.93	61.64	-96.53	39.27
Allen	-5.00	0.20	5.60	-6.37	2.30	6.15	27.34	1,048.66	9.78
Avoyelles	-10.30	3.50	15.35	-10.98	-3.02	13.46	6.63	-186.25	-12.33
Beauregard	-4.55	2.40	13.35	-8.01	0.45	6.74	75.94	-81.26	-49.54
Caddo	-10.60	0.90	17.60	-11.46	1.39	12.61	8.10	54.14	-28.34
Catahoula	-10.00	2.95	11.00	-6.17	-0.09	6.25	-38.25	-103.06	-43.22

Table 4.8. Continued.

Concordia	-12.55	4.95	16.50	-8.39	-3.70	6.76	-33.15	174.78	-59.02
East Carroll	-12.00	2.00	12.20	-9.13	1.42	10.56	-23.93	-28.85	-13.42
Franklin	-7.60	4.55	15.00	-12.55	-1.06	9.83	65.14	-123.24	-34.49
Madison	-8.35	2.50	12.70	-8.66	1.37	6.87	3.77	-45.32	-45.90
Ouachita	-5.00	2.40	11.20	-13.70	2.49	10.86	173.96	3.63	-3.05
Pointe Coupee	-12.95	0.60	20.00	-15.96	0.05	9.96	23.28	-92.04	-50.18
Rapides	-10.30	3.50	14.55	-9.28	-2.80	14.02	-9.95	-179.87	-3.67
Richland	-10.65	1.65	18.25	-8.32	1.28	7.40	-21.85	-22.27	-59.47
Tangipahoa	-7.60	0.00	15.80	-9.70	2.28	8.66	27.65		-45.22
Washington	-11.00	3.10	14.00	-7.75	0.20	10.73	-29.56	-93.60	-23.38
	Trend Stationary								
	Detrended TS			**Differenced TS**			**Percent Error (%)**		
Bossier	-13.21	-2.47	14.17	-10.20	0.00	13.75	-22.79	-100.00	-2.93
East Feliciana	-7.41	0.29	12.65	-12.70	3.65	14.05	71.39	1,153.65	11.04
Evangeline	-16.14	3.74	10.73	-10.00	0.70	13.45	-38.04	-81.26	25.36
Morehouse	-13.61	-0.55	10.78	-15.00	0.20	20.20	10.22	-136.56	87.30
Natchitoches	-10.78	-0.87	10.48	-10.90	0.80	20.00	1.14	-191.46	90.84
Red River	-8.93	-1.16	13.84	-11.00	0.00	16.10	23.14	-100.00	16.35
Saint Landry	-10.10	-1.86	9.25	-3.85	0.50	10.35	-61.87	-126.86	11.88
Tensas	-9.67	-2.21	10.20	-13.70	0.55	20.00	41.71	-124.86	96.14
West Carroll	-9.53	-1.79	11.31	-12.85	0.25	16.00	34.89	-113.95	41.43

Note: Differenced RW is obtained by taking the first differences of an I(1) yield series (the correct filter); Detrended RW are the residuals from removing the a simple linear trend from a RW yield series (the wrong filter); Detrended TS are the residuals from removing a simple linear trend from a TS yield series (the correct filter); Differenced TS is obtained by taking the first differences of a TS yield series (the wrong filter); and blank spaces are obtained because division by zero has no meaning.

4.3.3 Arkansas Soybean yields

Percentiles and percent errors from densities estimated from appropriately and inappropriately filtered soybean yields for Arkansas are presented in Table 4.9. Notice that in the first row and fourth column in Table 4.9 the effect of detrending a random walk yield series (the wrong filter) produces a probability density function with the following percent errors 32.60% (25^{th}), -49.43% (50^{th}), and -33.62% (75^{th}). County level yields show more sensitivity to inappropriate filtering. For instance, the highest impact as measured by the percent error is -447.89% (50^{th}) in Cross County, while the lowest is 0.89% (25^{th}) in Craighead County. Percent errors show clear patterns according to the type of erroneous transformation. For instance, an underestimation is produced in 80% of the cases as a result of detrending a random walk yield series (the inappropriate filter), while an overestimation is obtained in 75% of the cases as a consequence of first-differencing a trend-stationary series (the inappropriate filter). These two patterns are the same as the ones observed in Louisiana soybeans.

Table 4.9. Arkansas Soybean Yield Kernel Probability Estimates in Percentile Values.

County	**Random Walk**								
	Differenced R.W.			**Detrended R.W.**			**Percent Error (%)**		
	25th	**50th**	**75th**	**25th**	**50th**	**75th**	**25th**	**50th**	**75th**
Arkansas State Total	-2.00	1.00	3.50	-2.66	0.51	2.32	32.76	-49.43	-33.62
Arkansas	-2.40	0.65	3.00	-1.88	0.22	2.04	-21.77	-66.49	-32.10
Clark	-3.00	1.00	4.00	-2.86	0.75	3.21	-4.65	-24.92	-19.77
Clay	-2.90	0.00	4.00	-1.26	0.36	2.40	-56.62		-39.99
Craighead	-2.50	0.00	4.00	-2.48	0.06	2.86	-0.89		-28.58
Dallas	-4.00	0.00	4.00	-2.24	-0.19	2.45	-44.05		-38.64
Desha	-2.85	0.50	4.00	-3.59	-0.18	3.85	25.95	-135.90	-3.67
Faulkner	-3.80	1.00	5.00	-2.45	0.79	2.61	-35.62	-20.77	-47.74
Hempstead	-3.00	0.00	2.35	-1.47	0.50	2.15	-51.10		-8.58
Jefferson	-3.60	0.00	3.50	-2.62	1.23	2.76	-27.13		-21.19
Lee	-2.00	1.00	3.80	-2.66	0.45	2.09	32.89	-55.42	-45.03
Logan	-3.00	1.00	4.00	-2.00	-0.49	2.99	-33.27	-149.39	-25.27
Miller	-1.00	1.00	3.00	-2.26	-0.15	2.82	126.23	-115.50	-6.06
Mississippi	-4.50	1.00	6.00	-3.73	0.48	3.77	-17.21	-51.69	-37.16
Monroe	-2.55	1.00	3.05	-2.60	0.65	2.23	2.13	-34.58	-26.96
Nevada	-3.00	0.00	3.50	-2.35	-0.53	1.70	-21.55		-51.33
Perry	-2.00	1.00	3.30	-1.95	0.39	2.61	-2.27	-60.62	-20.84
Phillips	-2.00	1.00	4.00	-3.40	0.46	2.25	69.83	-54.02	-43.71
Prairie	-2.30	1.75	3.00	-2.68	0.61	2.41	16.34	-65.08	-19.69
Pulaski	-2.00	1.00	4.00	-1.94	0.68	1.95	-2.77	-32.40	-51.19
Randolph	-4.00	1.50	4.00	-1.87	0.32	3.31	-53.37	-78.99	-17.37
Saline	-3.00	0.00	4.00	-1.93	-0.70	2.49	-35.73		-37.74
Sebastian	-3.00	1.00	4.90	-3.51	-0.34	4.14	16.92	-133.94	-15.41
Yell	-2.00	0.00	2.10	-2.73	-0.33	2.27	36.31		8.23
	Trend Stationary								
	Detrended T.S.			**Differenced T.S.**			**Percent Error (%)**		
Ashley	-3.77	0.62	3.24	-3.00	0.10	5.00	-20.36	-83.94	54.26
Conway	-1.98	0.68	2.98	-3.50	0.50	3.50	76.98	-26.88	17.37
Cross	-2.53	0.18	2.68	-3.25	1.00	4.35	28.51	441.89	62.05
Independence	-2.38	0.64	2.19	-3.00	1.00	4.00	26.13	55.23	82.31
Jackson	-2.58	0.35	2.55	-3.00	1.00	5.00	16.49	186.48	95.82
Lawrence	-1.79	0.46	2.30	-2.90	2.00	4.25	61.78	333.63	84.70
Lonoke	-2.10	0.52	1.97	-2.00	0.00	4.00	-4.64	-100.00	102.84
White	-1.61	0.58	2.88	-3.00	0.25	4.00	86.29	-56.80	39.03

Note: Differenced RW is obtained by taking the first differences of an I(1) yield series (the correct filter); Detrended RW are the residuals from removing the a simple linear trend from a RW yield series (the wrong filter); Detrended TS are the residuals from removing a simple linear trend from a TS yield series (the correct filter); Differenced TS is obtained by taking the first differences of a TS yield series (the wrong filter); and blank spaces are obtained because division by zero has no meaning.

4.3.4 Louisiana Soybean yields

Percentiles and percent errors from densities estimated from appropriately and inappropriately filtered soybeans yields for Louisiana are presented in Table 4.10. The effect of detrending a random walk yield series (inappropriate filter) at the state level yields produce on average a probability density function with the following percent errors 0.22% (25th), 30.14% (50th), and -27.96% (75th). At the parish level, the greatest impact was 879.59% (50th) in Avoyelles Parish, while the lowest was -1.08% (25th) in Morehouse Parish. In this particular case, the percent errors reveal that an underestimation is dominant independently of the filter used. For instance, an underestimation is produced in 71% of the cases when detrending a random walk series, in 67% when differencing a trend-stationary series, and in 100% when differencing or detrending a stationary series. Only the underestimation case when detrending a random walk is the same as in Arkansas soybeans and Louisiana corn. As pointed out in the literature review in chapter 2 mechanical selection of a filter is not appropriate. While arbitrary detrending would reach correct probabilities in about 12% of the cases and 73% when differencing, these probabilities would be incorrect in 15% of the cases when yields are stationary.

Table 4.10. Louisiana Soybean Yield Kernel Probability Estimates in Percentile Values.

County	Random Walk								
	Differenced R.W.			Detrended R.W.			Percent Error (%)		
	25th	50th	75th	25th	50th	75th	25th	50th	75th
Louisiana State Total	-2.00	0.50	3.50	-2.00	0.65	2.52	0.22	30.14	-27.96
Acadia	-3.00	0.40	3.80	-2.78	-0.50	2.85	-7.25	-224.18	-24.89
Avoyelles	-2.75	0.05	3.80	-2.02	0.49	1.92	-26.47	879.59	-49.47
Bossier	-2.10	1.45	4.05	-4.06	0.12	3.20	93.55	-91.50	-21.04
Caldwell	-2.60	0.40	2.90	-3.85	0.55	3.44	48.15	36.52	18.77
Catahoula	-3.05	0.10	4.40	-2.13	0.25	2.36	-30.01	147.90	-46.32
Concordia	-3.95	-0.75	2.95	-4.00	1.02	2.42	1.32	-235.99	-18.04
East Carroll	-3.90	0.55	3.70	-4.34	1.60	4.74	11.20	190.77	28.12
Evangeline	-3.40	1.40	3.80	-2.52	-0.30	3.41	-25.80	-121.30	-10.30
Franklin	-3.70	0.95	3.90	-2.98	0.65	2.94	-19.40	-31.54	-24.54
Iberia	-3.25	0.55	3.90	-5.02	0.45	4.64	54.44	-17.38	19.03

Table 4.10. Continued.

Jefferson Davis	-4.00	0.65	4.10	-2.75	0.55	3.68	-31.35	-15.25	-10.19
Madison	-4.60	0.45	5.40	-2.58	0.14	3.24	-43.89	-69.41	-40.08
Morehouse	-3.85	0.55	4.00	-3.89	0.18	3.32	1.08	-66.67	-17.07
Natchitoches	-4.70	0.00	5.00	-3.61	0.61	3.55	-23.30		-29.02
Ouachita	-3.40	0.50	3.80	-3.78	0.57	2.18	11.17	14.86	-42.59
Pointe Coupee	-4.90	2.10	5.30	-2.63	-0.09	2.91	-46.37	-104.38	-45.15
Rapides	-2.95	0.10	4.45	-1.86	0.06	3.03	-37.03	-42.13	-31.97
Red River	-4.60	0.60	4.00	-3.31	-0.44	3.42	-28.02	-172.93	-14.45
Richland	-2.70	-0.30	3.00	-5.19	0.24	4.18	92.38	-180.30	39.35
Saint Landry	-2.80	1.00	4.15	-1.67	0.31	2.29	-40.43	-69.17	-44.76
Tensas	-2.95	0.10	5.55	-2.44	-0.27	3.28	-17.18	-374.77	-40.93
Vermilion	-3.60	0.70	3.90	-2.14	-0.03	1.94	-40.58	-103.63	-50.29
West Baton Rouge	-3.50	0.00	5.40	-2.69	0.39	3.59	-23.01		-33.48
West Carroll	-3.70	-0.80	4.30	-4.19	1.16	3.57	13.31	-244.80	-16.88
West Feliciana	-2.20	-0.10	4.40	-2.28	0.53	2.82	3.44	-626.05	-35.98
	Trend Stationary								
	Detrended T.S.			**Differenced T.S.**			**Percent Error (%)**		
Beauregard	-4.72	0.19	4.24	-1.80	0.30	3.60	-61.87	54.15	-15.03
Caddo	-3.43	-1.11	4.41	-2.95	0.00	2.70	-13.99	-112.25	-38.81
Iberville	-3.89	-0.98	4.22	-3.00	1.05	5.70	-22.81	-207.10	35.05
Saint Martin	-2.64	-0.07	2.71	-3.20	0.35	4.05	21.37	-567.58	49.68
	Stationary								
	Levels Stationary			**Detrended Levels**			**Percent Error (%)**		
Allen	20.10	22.40	25.80	-3.30	0.65	4.05	-116.42	-97.10	-84.30
Calcasieu	21.00	24.00	27.30	-3.60	-0.05	3.55	-117.14	-100.21	-87.00
Cameron	21.00	24.30	25.00	-4.00	-0.60	4.55	-119.05	-102.47	-81.80
Grant	22.90	26.50	31.50	-3.20	1.40	4.80	-113.97	-94.72	-84.76
Lafayette	23.50	27.80	30.70	-4.20	0.90	5.60	-117.87	-96.76	-81.76
	Stationary								
	Levels Stationary			**Differenced Levels**			**Percent Error (%)**		
Allen	20.10	22.40	25.80	-1.68	0.83	3.52	-108.36	-96.31	-86.37
Calcasieu	21.00	24.00	27.30	-2.69	0.78	3.23	-112.81	-96.76	-88.17
Cameron	21.00	24.30	25.00	-2.39	0.31	2.15	-111.36	-98.72	-91.42
Grant	22.90	26.50	31.50	-4.18	0.41	5.53	-118.25	-98.47	-82.43
Lafayette	23.50	27.80	30.70	-3.80	1.12	3.32	-116.18	-95.98	-89.20

Note: Differenced RW is obtained by taking the first differences of an I(1) yield series (the correct filter); Detrended RW are the residuals from removing the a simple linear trend from a RW yield series (the wrong filter); Detrended TS are the residuals from removing a simple linear trend from a TS yield series (the correct filter); Differenced TS is obtained by taking the first differences of a TS yield series (the wrong filter); Levels Stationary refers to the actual yields that are stationary (the correct filter is no transformation); Detrended Levels are the residuals from removing a simple linear trend from a ST yield series (the wrong filter); Differenced Levels is obtained by taking the first differences of a stationary yields (the wrong filter); and blank spaces are obtained because division by zero has no meaning.

5. SUMMARY IMPLICATIONS AND CONCLUSIONS

This study applied tests for unit-roots to historical crop yield data as an approach to identify how to filter the data prior to crop yield density estimation. Although tests for unit-roots have become standard methods in empirical time-series analysis, little emphasis has been given in the literature to applying these tests to identify empirically plausible models of crop yields (*i.e.,* stationary, simple linear trends, random walks). The first section of this chapter summarizes the main findings, followed by some insight into the implications of study in the second section. The third section outlines concluding remarks. There are limitations to the present study but suggestions for future research are discussed in the last section.

5.1 Summary

The main objective of this thesis was to assess the appropriateness of data filtering methods on crop yield density estimation with time-series data. One decision to be made in fitting a crop yield density function is whether to transform actual yields to a random process. The empirical literature often reports that crop yields must be transformed by linear-detrending, first-differencing or by a combination of first-differencing plus some sort of time-series model (*e.g.,* ARIMA models) prior to fitting a density function. Yet in other instances, the literature recommends that one general filter (*e.g.,* an ARIMA(0,1,2) mode) should be sufficient to transform all yields. This study relied on the use of unit-roots tests (*e.g.,* Enders, 1995) to determine whether yields need to be filtered, and if so, to identify the time series process adequate for the data at hand. It is important to emphasize that tests for unit roots have been available since the mid-1970s. What has been missing in applied work is a systematic use of such tests to identify nonstationary properties in yields which makes the choice of filters a trivial decision. This study filled that gap and outlined the steps needed to decide on appropriate filters.

It could be argued that the choice of filters is not as relevant as the final estimates of probabilities of reaching a certain yield. One way to address this concern is by calculating percentiles (*e.g.,* 25%, 50% and 75%) from crop yield densities estimated under the appropriate filter relative to some other, perhaps arbitrary, filter. One example that has practical relevance would be when a linear trend is removed from yield data prior to density estimation (call this nonparametric density F_1) versus the case when the unit-root tests suggest that yields should be filtered with first-differences to remove the unit-root prior to estimating the density (call this one F_2). A simple measure of the gap between these two empirical densities is a percent error = $[(F_1 - F_2)/F_2]*100$. This measure was calculated for all state and county level yields for corn and soybeans in Arkansas and Louisiana.

5.1.1 The Empirical Study

Augmented Dickey-Fuller tests of unit-roots were applied to annual corn yields for 30 counties in Arkansas and 25 parishes in Louisiana. For soybeans, there were 30 counties in Arkansas and 34 Parishes in Louisiana. The following hypotheses were tested. First, yields are stationary, meaning that there is no unit-root and that the data are random, possible around a constant mean; second, there is a unit-root in yields meaning that mean yields and variances are changing over the years; third, yields are trend stationary meaning that a simple linear trend can be used to model them; lastly, yields follow a unit-root process with a trend and possibly a constant. These hypothetical processes represent processes often reported in the empirical literature on crop density estimation. The results of the unit-root tests applied to the above 121 series suggested a dominance of yields that behave as random walks. In fact, random walks were adequate for 84% and 64% of the corn cases in Arkansas and Louisiana, respectively, and 71% and 73% of the soybeans cases in Arkansas and Louisiana, respectively. An often observed yield pattern is when yields grow at a constant rate from year-to year; this pattern can be easily accommodated by a simple linear trend. Trend-stationary yield series were adequate in 16% of Arkansas corn counties, 36% Louisiana corn parishes, 29% of Arkansas corn counties and 12% Louisiana soybean parishes. One surprising finding was that of random yields around a constant mean and variance. This process seemed appropriate in about 15% of soybean producing parishes in Louisiana. The above findings were used to trivially decide, as initially proposed, on the appropriate transformation (filter) to use for each crop yield series. For example, first-differencing yields that behave as random walks was the appropriate filter for most corn and soybean yields in Arkansas and Louisiana; clearly, linear detrending would be inappropriate in these cases. The remainder of the empirical analysis focused on the estimation of nonparametric kernel density functions of crop yields. The key question under investigation is the extent to which alternative crop yield density functions differ when calculated from appropriately filtered data (*i.e.,* unit-root tests and time series models) versus some other arbitrary method. A computation of percent errors between alternative probabilities was carried to reveal the impact of inappropriate filtering. The results revealed that the resulting percent errors can be negligible in some cases yet in others the percent errors can be sizeable. For example, errors in the magnitude of 908.12% were found for Arkansas corn, -1153.65% for Louisiana corn, -47.89% for Arkansas soybeans and -879.59% in Louisiana soybeans.

5.2 Implications for Agricultural Risk Management

Day (1965) was absolutely correct when he pointed out half-a-century ago that the science of crop yield density estimation was at the heart of the estimation of yield risk. The results of this thesis point towards improved crop yield probability density estimation and shed light on how to refine methods for calculating them. If risk assessments should rely on probability estimates from such density functions, then clearly crop risk management programs can benefit from methods that improve the estimation of risk *premia*. Risk management programs encompass crop insurance and revenue assistance among others and are designed on the premise of diminishing the risk faced by farmers. Farmers can freely decide on program participation, and for crop insurance, decide on the type and level of coverage. These decisions are based on risk preferences which are in turn based on the probability of a loss. However, as demonstrated in this thesis, the estimation of the probability that a yield loss will occur may differ with the choice of filters. As a matter of fact, an erroneous assessment of the true crop yield risk may result in issues such as unfair insurance *premia* that may decrease farmers' participation in the program. Other consequences of inaccurate yield probabilities may include in inefficient allocation of the government subsidies, indemnity support and revenue assistance programs. The two following subsections illustrate the importance of the findings in this study for two crop risk management programs, namely, a group risk plan (GRP) and average crop revenue election (ACRE).

5.2.1 Group Risk Plan

A farmer chooses to buy a crop insurance policy and decides on a coverage level in order to prevent a loss of revenue. The choice is based on an evaluation of the cost of the policy compared to the probability that a loss will occur. However, the cost of a policy may not be worth the true probability of the farmer's loss. Although a fair computation of the crop yield density is essential in rating crop insurance contracts, the estimation of the loss probability can be biased by using inappropriate data filtering which may affect farmers' program participation and also lead to loss-ratios that may jeopardize the sustainability of a program. For instance, let us suppose that a farmer producing corn in Logan County, Arkansas, buys a GRP insurance policy. Under this insurance plan, farmers select a coverage level between 70% and 90% of the long term county average yield and receive an indemnity if the actual crop county yield falls below the trigger yield (Edwards, 2009). Let us further assume that the farmer chooses an 85% coverage level. In order to illustrate the impact of inappropriate data filtering for this farmer, two probability density functions (PDF) have been estimated and are presented in Figure 5.1, after having rigorously applied unit-root testing and determined the adequate filter for corn yields in Logan County. The left-hand side PDF of Figure 5.1 has been generated based on appropriately filtered data (*e.g.,* first differencing a random walk) whereas the right-hand side PDF of Figure 5.1 has been estimated with inappropriately filtered data (*e.g.,* detrending a random walk). The long-term expected yield is 70.29 bu/acre (μ) and the trigger yield (TY) corresponding to an 85% coverage level (CL) is 59.74 bu/acre. Figure 5.1 shows that if the GRP contract is based on a distribution of appropriately transformed data, then the probability of falling below the trigger yield would be less than 0.25. On the contrary, if a GRP contract is based on a distribution misestimated due to inappropriately transformed data, then the probability of falling beneath the trigger yield is greater than 0.25. Consequently, it appears that the likelihood that a loss will occur would be different across alternative data approaches, and, in this particular case, inappropriate data filtering would lead to an overestimation of the probability of falling below the yield coverage. This would imply a premium that may not reflect and pay off the true risk faced by a producer.

Currently, corn and soybean farmers in Arkansas and Louisiana have experienced high insurance participation rates (RMA, 2010). Lately, there have also been significant annual fluctuations in the loss ratio (indemnity/premium) which continuity in the future may call the actuarial soundness of the Federal Crop Insurance Corporation (FCIC) programs into question;

thus, discouraging the use of current crop insurance products by farmers, insurance companies and government agencies. The empirical findings of this study give an opportunity to revise the methods used in calculating yield risk *premia*.

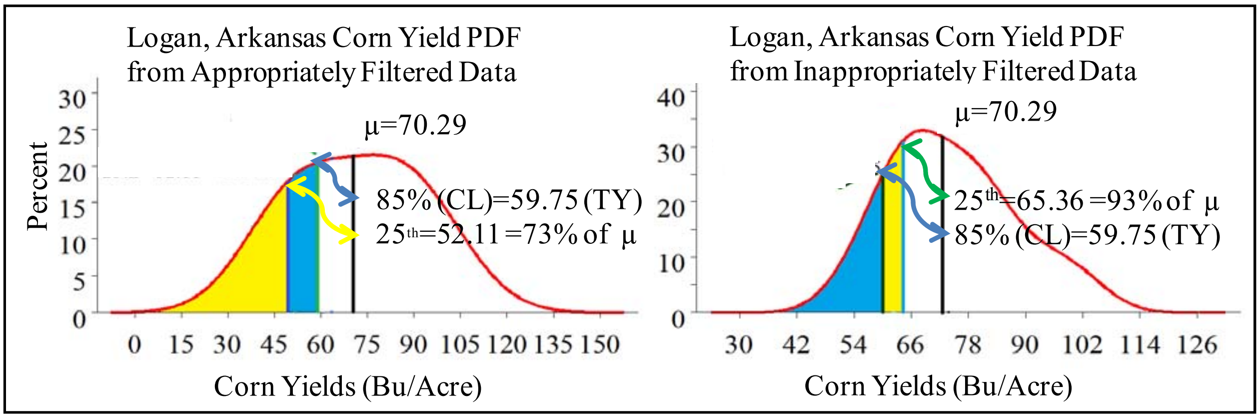

Figure 5.1. Alternative PDFs for Logan, Arkansas.

5.2.2 Average Crop Revenue Election Program

Average Crop Revenue Election (ACRE) is a new farm assistance program established under the Food, Conservation and Energy Act of 2008 with the objective of stabilizing agricultural risk. The eligible crops for this program include corn, cotton, soybeans, wheat, sorghum, barley, rice, oats, peanuts, dry peas, lentils, and chickpeas. Farmers enrolling in the ACRE program must sign up for all crops on the farm and the whole farm bill period (2009-2012), give up all countercyclical payments and 20% of their direct payments and accept a reduction of 30% of their marketing loans. The ACRE program will give right to a payout if the Corn Yields ACRE state revenue guarantee is greater than the Actual state revenue. The latter is the product of the state average yield (*e.g.*, average corn yield in 2009) and the season average price (*e.g.*, average corn price in the 2009/10 marketing year). The ACRE state revenue guarantee is 90% of the product of the ACRE price and the ACRE yield which is the simple average of the three more centered yields obtained during the last five years. Although these two important policy parameters include information on crop prices and yields, for the purpose of this research, only the latter is considered in order to highlight the implications of not using objective probability information in the design and evaluation of the ACRE program.

The point of interest is when the Actual yield falls below the ACRE yield. The five year Olympic average may or may not represent a farmer's actual yield since this method is not based on objective probability information which can be a major weakness of the ACRE program as "situations are rare in which… the decision maker has absolutely no subjective feelings or objective information about the probability distributions of outcomes…" (Barry, 1984 p. 32). For this reason, in this research, analyses of both types of yields (actual and average acre yields) using probability density functions were undertaken for Louisiana state corn yields. Actual yields are the state level average yields reported from the NASS from 1960 to 2008. The ACRE yield average series were built for each year from 1966 to 2008 using Olympic averages. For instance, the Louisiana corn state yields obtained in 1960-1964 were 27, 37, 28, 31 and 31 and the 1965 ACRE Olympic average yield is 30 bu/acre (*e.g.,* 28+31+31)/3). As unit-root testing suggested that the two series had a unit-root, first-differences were applied to achieve stationarity. A kernel density function was estimated for each case. The upper-PDF presented in Figure 5.2 corresponds to the ACRE Olympic yield averages, while the lower-PDF was estimated based on the actual yields. In the case of Louisiana corn yields, the inspection of Figure 5.2 shows that the probability of obtaining an actual yield greater than or equal to 97.29 bu/acre is 0.25 (green area of the lower-PDF of Figure 5.2), while the probability that the same yield occurs in the case of ACRE historical yield averages is clearly less than 0.25 (green area of the upper-PDF of Figure 5.2). On the other hand, there is a 0.25 probability that an ACRE Olympic yield average were equal to 87.39 bu/acre (the yellow area of the upper-PDF Figure 5.2), while there is a probability greater than 0.25 that the same yield will occur when analyzing the actual yields (blank area between the blue and brown lines plus the green area of the lower-PDF of Figure 5.2).

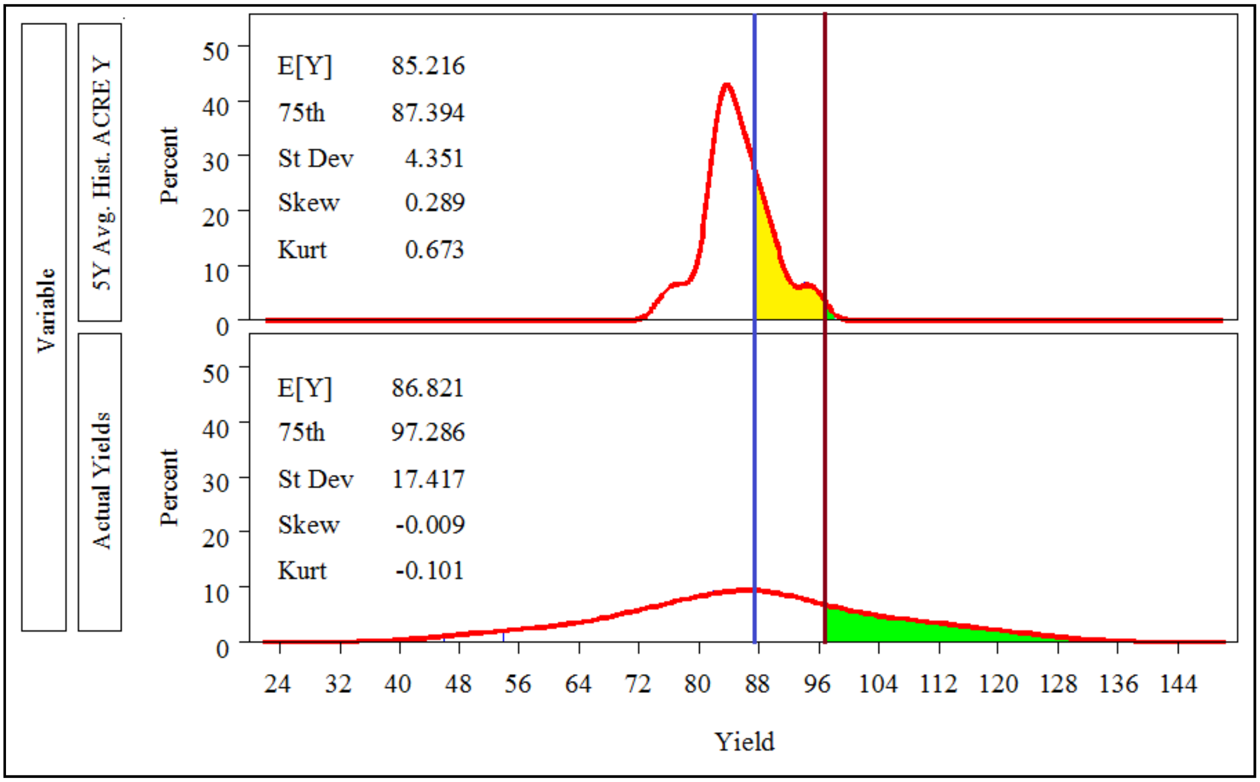

Figure 5.2. ACRE Program for Louisiana Corn.

In summary, the analysis of the ACRE program shows that there may be a higher probability mass associated with values equal to or smaller than the expected mean of the distribution function of Olympic averages than is the case for probabilities of yields obtained from aggregated historical records. Clearly, if the expected value of the distribution of Olympic averages is lower than the expected yield from historical records, and the probability space for the distribution of Olympic averages is contained in the probability space of expected yields from historical records, then, the expected compensation under yield risk associated with the ACRE program would be lower. The lower expected payout may deter some from ACRE program participation.

5.3 Conclusions

This study has shown that when fitting empirical yield density functions, the choice of filters used to transform yields prior to density estimation can be made simple through the application of unit-root tests. Although these statistical tests have been available for some time, guidance is provided on their implication to crop yield density estimation. The empirical evidence points towards crop yields being characterized by a number of different processes, implying that a single filtering technique cannot be generalized to all crops and geographic areas, a finding consistent with that of Zapata and Rambaldi (1989). These findings lead to the useful conclusion that the existing filtering methods (*e.g.,* ARIMA (0,1,2)), suggested by previous works, would be appropriate for approximately 74% of the studied counties in Arkansas and Louisiana. However, for the 26% remaining cases, the same filter would produce unreliable crop yield densities. Furthermore, detrending historical crop yields (the most employed filter) would be justified in only 22% of the cases, while for the remaining 4% are already stationary, and thus no filter is needed. The empirical results also revealed discrepancies (*i.e.,* sizeable percent errors) between alternative probability estimates as a consequence of inappropriate filtering.

Since most of the crop risk management programs rely on probability density estimation, it was pertinent to link the results of this research to some of these programs. In particular, an evaluation of the Group Insurance Plan (GRP) has been analyzed in order to illustrate the impact of alternative crop yield data filtering on probability estimates. If crop yield series behave as a random walk, then detrending will result in an overestimation of the likelihood of achieving a certain yield level. In other words, the application of inappropriate data filters may overestimate the probability loss which may result in higher risk premium estimates, lowering the chance of receiving indemnity payments. Consequently, farmers' participation rates in the programs will tend to decrease impacting the sustainability of crop insurance programs. In the case of the ACRE program, an Olympic average may not be representative of historical yield records, and if so, the likelihood of a payout decreases.

5.4 Limitations and Further Research

The main interest of this study was to demonstrate the usefulness of unit-root tests in identifying time-series properties of crop yield data and how such information can be used in filtering yield data prior to density estimation. Plainly stated, in choosing a filter opt for the one that is more consistent with time-series properties of yields. No effort was made to identify the best time-series model for the filtered series or their relationship to other factors, such as rainfall and temperature, which may influence yields. Future work should revise the probability estimates reported in light of those improved models.

The findings in this study are empirical using aggregated historical data from 1960-2008. Because unit-root tests of the augmented Dickey-Fuller type are more reliable in larger samples, one useful extension of the study would be to design a Monte Carlo experiment based on the data-based findings reported here under various sample sizes, say 25, 50, 75, 100, and 200 observations. The experiment may allow for simulation of a variety of models found in this evaluation of corn and soybean data for Arkansas and Louisiana, including random walks, ARIMA models of various orders, simple linear trends, and stationary series. An added dimension of the experiment would be its formulation in a panel data framework as recently done in the literature (*e.g.,* Atwood, Shaik and Watts, 2003). Drawing from the extensive existing econometric literature on the robustness of unit-root tests with small samples, and the preliminary analysis in Zapata and Rambaldi (1989), such an experiment would shed useful light on the performance of nonparametric crop yield density estimators with nonstationary data and on the direction the science of crop yield density estimation should continue.

Another natural extension of this work is an extension of crop yield density estimation for other crops and feedstocks that can be used in biofuel production. Empirical evidence on feedstock yield risk would be valuable in assessing the commercial viability of alternative biofuel technologies and such information is lacking in the Delta region.

REFERENCES

Akaike, Hirotugu. "A new look at the statistical model identification." *IEEE Transactions on Automatic Control*, 19(June, 1974):716–723.

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

Barry, P.J. *Risk Management in Agriculture.* Iowa State University Press, 1984.

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

Day, R.H. "Probability Distributions of Field Crop Yields." *Journal of Farm Economics*, 47,3(August 1965):713-41.

Dickey, D.A. and W.A. Fuller. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association*, (1979):427–431.

Edwards, W. "Group Risk Plan (GRP) and Group Risk Income Protection (GRIP)" http://www.extension.iastate.edu/agdm/crops/html/a1-58.html, Accessed 2010, March 31

Enders, W. *Applied Econometric Time Series*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons, 1995.

Foote, R. J., and L. H. Bean. "Are Yearly Variations in Crop Yields Really Random?" *Journal of Agricultural Economic Resources*, 3(1951):23-30.

Gallagher, P. "U.S. Corn Yield Capacity and Probability: Estimation and Forecasting with Nonsymmetric Disturbances." *North Central Journal of Agricultural Economics*, 8,1(January 1986:)109-22.

____. "U.S. Soybean yields: Estimation and Forecasting with Nonsymmetric Disturbances." *American Journal of Agricultural Economics*, 69,4(November 1987):796-803.

Goodwin, B. K., and A.P. Ker. "Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts." *American Journal of Agricultural Economics*, 80,1(February 1998):139-53.

Gommes, R. "Non-Parametric Estimation of Crop Yield Forecasting, a Dictating Case Study for Zimbabwe." Paper presented at the EU/JRC meeting on Remote Sensing Support to Crop Yield Forecast and Area Estimates, 30 Nov. - 1 Dec. 2006, Stresa, Italy.

Hamilton, J.D. *Models of Nonstationary Time Series.* Time Series Analysis. Princeton, NJ: Princeton University Press, 1994.

Hill, R.C., W.E. Griffiths and G.C. Lim. *Principles of Econometrics Third Edition*. Hoboken, NJ: John Wiley & Sons, Inc, p579, 2008.

Houck, J.P. and P.W. Gallagher. "The Price Responsiveness of U.S. Corn Yields." *American Journal of Agricultural Economics*, 58,4(November 1976):731-734.

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

Kaylen, M.S., and S.S. Koroma. "Trend, Weather Variables, and the Distribution of U.S. Corn Yields." *Review of Agricultural Economics*, 13,2(July 1991):249-58.

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

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

Luttrell, C.B., and R.A. Gilbert. "Crop Yields: Random, Cyclical, or Bunchy?" *American Journal of Agricultural Economics*, 58,3(August 1976):521-31.

Moss, C.B., and J.S. Shonkwiler. "Estimating Yield Distributions with a Stochastic Trend and Nonnormal Errors." *American Journal of Agricultural Economics*, 75,4(November 1993):1056-62.

Menz, K. M. and P. Pardey. 1983. "Technology and U.S. Corn Yields: Plateaus and Price Responsiveness." *American Journal of Agricultural Economics*, 65(3):558-62.

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

Nelson, C.H., and P.V. Preckel. "The Conditional Beta Distribution as a Stochastic Production Function." *American Journal of Agricultural Economics*, 71,2(May 1989):370-78.

Norwood, B., M.C. Roberts, and J.L. Lusk "Ranking Crop Yield Models Using out-of-Sample Likelihood Functions." *American Journal of Agricultural Economics*, 86,4(2004):1032-43.

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

Ramirez, O.A., S. Misra, and J. Field. "Crop-Yield Distributions Revisited." *American Journal of Agricultural Economics*, 85(February 2003):108-20.

Risk Management Agency Website (RMA). www.rma.usda.gov. Accessed 2010, March 31.

Phillips P.C. and P. Perron. “Testing for a Unit Root in Time Series Regression.” *Biometrika*, 75,2(June, 1988):335-346.

Sherrick, B.J., F.C. Zanini, G.D. Schnitkey, and S.H. Irwin. “Crop Insurance Valuation under Alternative Yield Distributions.” *American Journal of Agricultural Economics*, 86,2(May 2004):406-19.

Silverman, B. W. *Density Estimation*, New York: Chapman & Hall, 1986.

Taylor, C.R. “Two Practical Procedures for Estimating Multivariate Nonnormal Probability Density Functions.” *American Journal of Agricultural Economics*, 72(February 1990):210-17.

Thompson, L.M. “Weather and Technology in the Production of Corn in the U.S. Corn Belt.” *Agronomy Journal,* 61(1969):453-456.

Turvey, C.G., and J. Zhao. “Parametric and Non-Parametric Crop Yield Distributions and Their Effects on All-Risk Crop Insurance Premiums.” Working Paper WP 99/05, Department of Food, Agricultural and Resource Economics, University of Guelph, Guelph, January 1999.

Zapata, H.O., and A.N. Rambaldi. “Effects of Data Transformation on Stochastic Properties of Economic Data,” Research Report No. 682. Department of Agricultural Economics, Louisiana State University, Baton Rouge, 1989.

Zapata, H.O., and A.N. Rambaldi. “Effects of Data Transformation on Stochastic Properties of Economic Data,” Paper presented at the annual meetings of the American Agricultural Economics Association, Baton Rouge, Louisiana, 1989.

VITA

David Isaias Maradiaga Pineda was born in Tegucigalpa, Honduras. After completing high school in 2001 at the Instituto Departamental de Oriente, he matriculated in Escuela Agricola Panamericana (El Zamorano) in 2002, where he completed his Bachelor of Science in agriculture with a minor in food science in 2005. He enrolled in Louisiana State University in 2008 as a graduate student in the Master of Science program in agricultural economics in the Department of Agricultural Economics and Agribusiness. Currently, he is looking forward to enrolling in a doctoral program.