ESSAYS ON INCORPORATING RISK MODELING TECHNIQUES IN AGRICULTURE

A Dissertation

by

RYAN A. LARSEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2011

Major Subject: Agricultural Economics

Essays on Incorporating Risk Modeling Techniques in Agriculture

ESSAYS ON INCORPORATING RISK MODELING TECHNIQUES IN AGRICULTURE

A Dissertation

by

RYAN A. LARSEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	David J. Leatham
	Dmitry V. Vedenov
Committee Members,	James Mjelde
	Clair J. Nixon
Head of Department,	John P. Nichols

August 2011

Major Subject: Agricultural Economics

ABSTRACT

Essays on Incorporating Risk Modeling Techniques in Agriculture. (August 2011)

Ryan Larsen, B.A., Idaho State University; M.S., University of Idaho

Co-Chairs of Advisory Committee: Dr. David J. Leatham
Dr. Dmitry V. Vedenov

Measuring, modeling, and managing risk has always been an important task for researchers. Many of the traditional assumptions relied on in risk research, such as the assumption of normality and single period optimization, have proven too restrictive and alternative risk management tools have been developed. The objective of this dissertation is to explore and apply these tools to analyze geographical diversification. The first step to analyze geographical diversification is to understand how different climate and spatial variables impact yields. Yield dependencies for wheat, cotton, and sorghum are estimated using linear correlation and copulas. The copulas provide an alternative to linear correlation. The results of the different dependency estimations indicate that there is a significant difference between the results.

The next step is to analyze geographical diversification in a portfolio setting. Traditional portfolio optimization has assumed that risk and dependence are symmetric. Using a single period model, an asymmetric risk measure, conditional value at risk, and asymmetric dependence measure, copulas, are implemented into the portfolio optimization model. The efficient frontiers under both symmetric and asymmetric assumptions show that ignoring the asymmetric nature of the data could lead to optimal

portfolio allocations that could underestimate the actual risk exposure. The implication of these results provides researchers with motivation to move beyond the standard assumptions of linear correlation and normality.

Building on the single period problem, a multi-period portfolio model is formulated using discrete stochastic programming. One key in formulating a discrete stochastic program is the representation of uncertainty. Scenario generation is a method to obtain a discrete set of outcomes for the random variables. A moment matching routine is developed to capture the first four moments of the variables and the multivariate relationship is modeled using copulas. The results show that the moment matching routine closely captures the higher moments of the data. The results also indicate that there are possible gains from geographical diversification. Wealth levels increased for all three regions when production is diversified over the different regions. The optimal land allocation was dependent upon the base acreage assumption.

ACKNOWLEDGEMENTS

I would like to thank my committee chairs, Dr. Leatham and Dr. Vedenov, and my committee members, Dr. Mjelde, and Dr. Nixon, for their guidance and support throughout the course of this research. I would also like to thank NOAA for providing part of the funding necessary to complete this research.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience. I would also like to thank my dear wife, McKenzie, for enduring through the many years of my schooling. I would like to thank Lilly, Maggie, and Sophia for always providing me with a true perspective of what is important. Finally, thanks to my mother and father for instilling in me a desire for furthering my education.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

Innovations in both statistics and finance have led to the development of new methods to analyze risk management decisions. Specifically, the advances in measuring dependency to model joint distributions and the development of alternative risk criterions has allowed both practitioners and researchers to look at old problems in a new way (Clemen, and Reilly, 1999). Recently, these new methods of incorporating alternative dependency structures and risk measures made their way in the agricultural economics research literature (Ozaki, Goodwin, and Shirota, 2008; Vedenov, 2008; Zhu, Ghosh, and Goodwin, 2008). There are limited applied applications in agricultural settings, thus, there is a need for research to further investigate these new methods and compare them with the traditional methods.

Traditional risk management has relied heavily upon the mean-variance portfolio model developed in 1952 (Markowitz, 1952). The implementation of Markowitz's Portfolio Theory marked one of the great beginnings of financial risk management (Elton, and Gruber, 1997). Building on the work done by Markowitz, others such as Tobin and Sharpe have enhanced the applications of portfolio theory (Sharpe, 1963; Tobin, 1958). Agricultural economists approach to the portfolio problem mirrored that of Markowitz (Heady, 1952; Peterson, Schurle, and Langemeier, 2005; Robison, and Brake, 1979).

This dissertation follows the style of *American Journal of Agricultural Economics.*

Portfolio theory relies on a few key assumptions. The first is that linear correlation captures the dependence structure of the data. The second is that returns are normally distributed. This assumption of multivariate normality implies that the random outcomes can be fully described by the marginal distributions and the dependency structure is characterized by the correlation coefficient (Ane, and Kharoubi, 2003). The assumption of multivariate normality also implies that variance is the appropriate risk measure. The problem encountered in working with real world data is that often there exists both skewness and leptokurtosis in the data. This has led researchers to develop alternative probabilistic models (Krokhmal, 2007). These new probabilistic models have focused on working with univariate data problems. The implementation of alternative multivariate distributions relies heavily on the previously mentioned univariate models. This presents two problems. The first is that the marginal distributions of the multivariate vector all belong to the same univariate distribution. The second problem is that the dependence measure is taken from the marginal distributions (Embrechts, McNeil, and Strauman, 2002). Thus, there is a need for a more flexible approach to modeling multivariate data.

One method that has been used in the finance and insurance industries is that of the copula function (Accioly, and Chiyoshi, 2004; Clemen, and Reilly, 1999; Patton, 2002). The copula function links the n-dimensional distribution function to their one-dimensional margins. The function is continuous and characterizes the model's dependence structure. One of the key advantages of using the copula is that the appropriate model used for the specifying the dependence structure between the variables can be selected independent of the choice of the marginal distributions. In other words, without imposing any assumption on the marginal distributions, the dependency between the variables can be estimated. Utilizing this copula function, risk management routines such as portfolio optimization can incorporate these alternative marginal distributions and dependency structures.

Another issue of vital importance in risk management is the development of coherent risk measures (Acerbi, 2002; Acerbi, 2007; Alexander, and Baptista, 2004; Artzner, et al., 1999; Buch, and Dorfleitner, 2008; Csóka, Herings, and Kóczy, 2007). Coherent risk measures have been defined as risk measures that satisfy the properties of monotonicity, convexity, and translational invariance. Examples of coherent risk measures used in risk research are the LPM (Lower Partial Moments) and CVaR (Conditional-Value-at-Risk) approach. Both measures have been shown to exhibit the properties of a coherent risk measure (Alexander, and Baptista, 2003). The application of these risk measures are rare in the agricultural economics literature (Zylstra, Kilmer, and Urysasev, 2003). These authors used the LPM approach to measure the effectiveness of a dynamic hedging strategy of feeder cattle. Thus, there is a need for a better understanding of the properties and applications of coherent risk measures in agriculture.

The objectives of this dissertation are to implement recent innovations in statistics and finance to help analyze risk management issues in agriculture. The first objective is to better understand geographical diversification by analyzing the spatial relationships between yields. Geographical diversification has been researched on a limited basis (Just, and Pope, 2003). Geographical diversification consists of spreading production areas into regions that reduce yield and price risk. Technological advances in both production and communication have allowed producers to cross county and state lines to diversify their production portfolio. Little research has been done concerning geographic diversification and the problem faced by many firms is that they have minimal information to determine the "best" place to geographically diversify operational activities and accurately measure the risk involved with this type of diversification. Specifically, a tradeoff equation is estimated to help determine the relationship between yield dependence, elevation, and climate data. The yield dependence parameter is measured in three ways, Pearson correlation, rank correlation using copulas, and tail dependence coefficients.

The second objective is to apply coherent risk measures to a production agricultural problem. This is addressed by analyzing geographical diversification at the enterprise level. A representative farm is modeled. Conditional-Value-at-Risk (CVaR) is used to measure the risk of geographical diversification. A portfolio optimization algorithm is developed using CVaR as the risk criterion. The economic implications of using CVaR as a portfolio constraint are discussed. This is done by comparing the results and the EV frontier when using the CVaR and the standard mean-variance approach.

The last objective is to develop a discrete multi-period stochastic program to optimize a farmer's portfolio, where the portfolio is the acreage allocations between three distinct geographic areas. The farmer has the ability to allocate wheat production acreage over three dry land production regions, Montana, Colorado, and Texas. The portfolio allocation decisions take place at discrete time points. At these discrete points the farmer has to evaluate the previous year's market conditions and the composition of the enterprise diversification. At the same time, the farmer must evaluate future conditions such as expected future yields and prices. All this information is then used by the farmer to reallocate or adjust the land allocation over the different production regions. This may involve increased short term or long term borrowing because of increased operating expenses, machinery purchases, and land purchases or the retirement of debt in good income years. This same decision process continues through the time periods of the model.

The results of the discrete multi-period stochastic optimization are presented. The optimization algorithm consists of maximizing expected utility of wealth by allocating acreage levels in the three different regions. A dynamic analysis of the optimal acreage allocations over time is estimated as well as how these allocations change with different levels of risk aversion. These results and modeling technique provide a foundation for expanding the standard single period portfolio optimization problem to multi-periods. This modeling technique also provides a framework to analyze other farm financial decisions, farm growth decisions, and even could be applied to loan portfolios from a lenders perspective

The remainder of this dissertation is organized as follows. The relationship between yield dependencies and spatial factors is discussed in Section 2. This analysis is based on data collected from wheat, sorghum, and cotton counties. Historical yield data were collected for each county. In addition, geographic data such as elevation and latitude/longitude were collected for each location. These variables are used to analyze the sensitivity of the relationship between areas based on these geographic parameters.

Enterprise level geographic diversification is discussed in Section 3. Three possible production locations, one in Texas, one in Colorado, and one in Montana are modeled to simulate the effect of geographical diversification on a given enterprise. Alternative dependency structures are used in the risk estimation based on copulas. Conditional-value-at-risk (CVaR) is used as the risk measure for this analysis. CVaR is calculated using both the multivariate normal approach as well as the copula based approach. The final essay is discussed in Section 4. This essay applies a discrete stochastic programming to the issue of geographic diversification. It extends the single period portfolio model to multiple periods. Section 5 is the final section and the results and implications of this research are discussed. In addition, future research opportunities are discussed.

2. MODELING ALTERNATIVE DEPENDENCE STRUCTURES: AN APPLIED CASE TO WHEAT, COTTON, AND SORGHUM

2.1 Introduction

Agriculture is inherently risky facing production, market, and cost risks (Boehlje, and Lins, 1998; Escalante, and Barry, 2001; Hardaker, et al., 1997; Harwood, et al., 1999; Just, and Pope, 2002; Moschini, and Hennessy, 2001; Turvey, Driver, and Baker, 1988). Mishra and Lence (2005 p.131) defined risk "… as the uncertainty faced by a firm (be it an individual, agribusiness, or lender) that affects its welfare." To help manage risk one must understand the degree of dependence among the different management options. (Measuring the degree of dependence is discussed later.) Mishra and Lence (2005) classify risk management strategies into two categories: within-firm and risk-sharing strategies. Within-firm strategies include enterprise diversification, reducing leverage, gathering additional information about future scenarios, and increases in liquidity. Risk-sharing strategies consist of insurance, off-farm income, and the use of contracts such as futures and options.

Risk management strategies utilized by producers tend to vary by size and composition of the agricultural entity (Mishra, and Lence, 2005; Pope, and Prescott, 1980). An industry that was once composed mainly of family farms has now been segmented into three sizes: large-industrial companies, commercial-scale family operations, and the traditional small family farm (Featherstone, et al., 2005). Large-industrialized companies are diversifying risk through vertical integration and multinational operations (Boehlje, and Lins, 1998; Handy, and MacDonald, 1989). Commercial-scale family operations utilize risk management tools such as hedging, insurance, and crop diversification (Mishra, and El-Osta, 2002). Small scale family farms are diversifying by depending on off-farm income (Harwood, et al., 1999; USDA, 2000).

Another risk management strategy that is starting to be employed by multi-national and commercial-scale farms is diversifying their portfolio geographically. Production operations are geographically separated to reduce production risk and/or locate operations closer to processing plants to reduce market risk (Davis, et al., 1997). A limited number of studies have addressed farm level effects of geographic diversification (Davis, et al., 1997; Krueger, et al., 1999; Nartea, and Barry, 1994). Additional studies are necessary concerning geographical diversification to more fully understand the ramifications of such management strategies. Two problems faced by producers considering geographical diversification are what location(s) "best" diversifies their risk and what are the additional costs associated with geographical diversification.

One of the keys to answering these questions is to understand how yields are related (dependent) in different regions. Historically, the Pearson correlation coefficient has been relied upon to measure the dependence between two variables. This coefficient captures the linear dependence if the variables are normally distributed (Accioly, and Chiyoshi, 2004), but often, the normality assumption is not valid (Just, and Weninger, 1999). To capture the dependence between two variables without having to force the assumption of normality, copulas have begun to be used (Clemen and Reilly, 1999).

Copulas allow the flexibility of modeling the marginal distributions independently of the joint distribution. The estimated copula then 'couples' the individual marginal distributions to the respective joint distribution. As with most flexible measures, copulas have additional complexity (costs) associated with estimating them relative to Pearson's correlation coefficient.

Providing information on how various factors influence changes in yield dependencies associated with geographical diversification is the primary objective of this study. To obtain this objective, yield dependence functions based on changes in elevation, temperature, and precipitation are estimated. A secondary objective is to analyze the impact of using alternative dependency measures. As such, yield dependency functions are estimated based on different dependency measures: Pearson's correlation and various copula forms. Using real world data does not allow for the determination of the "best" measure, but does allow for comparison of the different measures. This research not only extends the current literature on geographical diversification by taking a more detailed examination of factors impacting yield dependence, but also extends the copula literature by comparing estimation results using linear correlation and copulas. The impact of using alternative dependence measures has only recently begun to be examined.

2.2 Literature Review

Managing risk by geographical diversification is not new; it has been used frequently in the banking and real estate industries. Liang and Rhoades (1988) using the changes in banking regulations in the late 80's as motivation, studied the impact of geographical diversification in the banking industry. Because of regulation changes, which allowed banks to expand into different regions, banks had begun to expand beyond state borders. Ling and Rhoades (1988) found that geographic diversification reduces insolvency risk, but increases in operating risks may occur because of increased management and acquisition costs. In another banking study, the impact of geographic diversification was specifically applied to small banks that were acquired by larger banks (Rose, and Wolken, 1990). Mergers appeared to provide no long run advantages for the small banks. In the short run, mergers, however, provided opportunities for entry into new markets.

Ehling and Ramos (2006) examined differences between sector and geographic diversification using industries within the Euro zone. They argued that with the implementation of the Euro, gains associated with geographic diversification are diminished. Using a mean-variance efficiency test, Basak et al. (2002) tested whether companies are better off by sector or geographical diversification. Results depend on the constraints imposed on the model. If short-selling constraints are imposed, geographic diversification outperforms sector diversification, whereas, the two strategies are statistically equivalent if the problem is unconstrained. Kim and Mathur (2008) suggested geographical diversification increases operating costs but also increases return on equity and return on assets when compared to industrially diversified firms. Results from previous studies in non-agriculture sectors suggest there are possible gains from geographic diversification.

Within an agricultural setting, results of studies on geographical diversification are inconsistent. Kreuger et al. (1999) showed that a grape grower could increase profits and decrease risk by producing in the U.S. and Chile. Nartea and Barry (1994) concluded that there are no realizable increases in net returns for individual grain growers from diversifying geographically in central Illinois. Costs associated with geographical diversification included in their model were transportation costs, monitoring costs, and losses because of poor machinery coordination. Davis et al. (1997) found an inverse relationship between Georgia peach orchards yield correlations and distance between the two orchards. Using farm level data, they estimated the volatility in yields that could be reduced by operations using spatially diverse orchards. Davis et al. (1997) concluded that correlation between yields is reduced by 2.28% for every additional mile the orchards are separated.

Absent from these studies was a detailed discussion concerning the measurement of the dependence between yields. The development of copulas has allowed for a new examination of dependence structure between variables. Copulas are a statistical tool used to model multivariate relationships. Although, copulas were developed in the 1950's, they have just recently been incorporated into finance, statistics, economics, and agricultural research (Accioly, and Chiyoshi, 2004; Ane, and Kharoubi, 2003; Bai, and Sun, 2007; Clemen, and Reilly, 1999; Dowd, 2005; Embrechts, McNeil, and Strauman, 2002; Genest, and Favre, 2007; Hennessy, and Lapan, 2002; Patton, 2002). The use of copulas in the agricultural literature is limited. Vedenov (2008) found that the flexibility allowed by copulas provided an efficient method for estimated the joint distributions for crop yields. Power and Vedenov (2008) model a dynamic hedge ratio using a copula-GARCH. They conclude the copula-GARCH methodology led to better results than the standard GARCH. Copulas have also been used in designing whole farm insurance (Zhu, Ghosh, and Goodwin, 2008). Zhu, Ghosh, and Goodwin (2008) use copulas to model the relationship between corn and soybean prices. The implementation of copulas also allowed the authors to individually model the marginal distributions of the crops separately. They found this approach provides a better fit than using the same marginal distribution for each crop.

2.3 Data

Data used consists of geographical, climate, and historical county yield data for non-irrigated wheat, cotton, and sorghum. Geographical and climate data consists of average temperature, precipitation, and elevation. The length of the yield data series varies slightly for each commodity. The largest block of continuous yield data available at the time of data collection were used. To be included, a county must have more than 10,000 harvested acres based on 2006 crop production crop. Ten-thousand acres is arbitrary chosen, but represents an area that indicates in most cases more than one producer.

2.3.1 Yield Data

County level wheat yields are for 1976 to 2001 for wheat (both spring and fall). More recent yields were not included because of lack of data for many of the counties at the time of data collection (USDA-NASS, 2008). To be included, a county had to meet the following criterion; First, the analysis is limited to the major wheat production regions of the U.S.; therefore, only counties in the following states are included, Colorado, Idaho, Kansas, Montana, Nebraska, North Dakota, Oregon, Oklahoma, South Dakota, Texas, Utah, or Washington (Figure A.1). Unfortunately, many Oklahoma counties could not be included because of large gaps of missing data. Three hundred and eighty counties met these two criteria (Table A.1).

Sorghum county yields are from 1978 to 2001. States included in the analysis are Kansas, Oklahoma, South Dakota, New Mexico, Nebraska, Colorado, and Texas (Figure A.1). Fifty counties are included with the non-irrigated sorghum data set (Table A.1). Cotton yields are for years 1977 to 2001 for upland cotton (Table A.1). Thirty-nine counties are included in the analysis from five states: Arkansas, Louisiana, Mississippi, Oklahoma, and Texas (Figure A.1).

2.3.2 Geographical Data

To provide a consistent location across the counties for geographical and climate data, the county seats are used to represent each county. Elevation for each of the county seats are obtained from the website Lat-Long.com. Thirty year average annual temperature and precipitation for weather stations located at or near the county seats are used to represent climate conditions (USDA-NRCS, 2008). Preliminary estimations included latitude and longitude variables. Latitude and longitude variables are highly collinear with the other variables and were dropped from the analysis.

2.4 Modeling Dependence

Modeling dependence is often overlooked and sometimes misunderstood. As previously noted, the Pearson linear correlation is usually considered to be the dependence measure without considering other measures. In fact, linear correlation is the measure of dependence forming the basis of different finance theories, including the Capital Asset Pricing Model, the Arbitrate Pricing Theory, and Markowitz Portfolio Theory (Embrechts, McNeil, and Strauman, 2002). These theories assume multivariate normally distributed returns. The assumption of normality, however, may not be appropriate. In the case of non-normal distributed variables, other measures of dependence may be more appropriate.

Embrechts, McNeil, and Strauman (2002, p.15) define quantification of the relationship between two random variables as summarizing "… the dependence structure between two random variables in one number." They identified five desirable properties of dependence between two random variables, x and y. Let the dependence between x and y be measured by δ, the five properties are:

1. Symmetry: $\delta(x, y) = \delta(y, x)$
2. Normalization: $-1 \leq \delta(x, y) \leq 1$
3. Comonotonic and Countermonotonic

$$\delta(x, y) = 1 \Leftrightarrow x, y\ comonotonic$$

$$\delta(x, y) = -1 \Leftrightarrow x, y\ countermonotonic$$

4. For T: $\mathbb{R} \rightarrow \mathbb{R}$ strictly monotonic on the range of

$$\delta(T(x), y) = \begin{cases} \delta(x, y) & T\ increasing \\ -\delta(x, y) & T\ decreasing \end{cases}$$

5. Independence

$$\delta(x, y) = 0 \Leftrightarrow x, y\ independent$$

Property 1 states that the measure of dependence (δ) between two random variables is equal regardless of variable ordering. In other words, the dependence between x and y is equal to the dependence between y and x. Property 2 states the measure of dependence should fall within the range -1 to 1. This enables comparisons to be made across different dependence measures. Property 3 (comonotonic) states that if the dependence between x and y is equal to 1, then x and y are considered to move perfectly together. Conversely, property 3 (countermonotonic) states that if the dependence between x and y is equal to -1, then x and y will move exactly opposite. Property 4 reflects the increasing nature of positive dependence and decreasing nature of negative dependence. Finally, property 5 states if the dependence between the two variables is equal to zero, x and y are independent. In a perfect world, a dependence measure would satisfy all five of these properties.

2.4.1 Pearson's Correlation Coefficient

Pearson's correlation coefficient only satisfies the first two properties(Embrechts, McNeil, and Strauman, 2002). This correlation coefficient is defined as

$$\rho_{xy} = \frac{cov(x,y)}{\sqrt{var(x)var(y)}} \tag{2.1}$$

where ρ_{xy} is a measure of linear dependence which takes on values between -1 and 1 (Wackerly, Mendenhall III, and Scheaffer, 2002). When $\rho_{xy} = 0$, x and y are said to be independent. If $\rho_{xy} = 1$ or $\rho_{xy} = -1$, then x and y are perfectly linearly dependent. As noted earlier, one potential problem with the Pearson's measure of dependency is that it is only appropriate if the data are considered to be multivariate normal (Embrechts, McNeil, and Strauman, 2002). Multivariate normal distributions are fully described by a mean vector and covariance matrix. If the data is not multivariate normal, then linear correlation may not accurately measure the dependence between variables.

2.4.2 Rank Correlation

An alternative to linear correlation is rank correlation. Rank correlations depend solely on the copula of the bivariate distribution and not on the marginal distributions (McNeil, Frey, and Embrechts, 2005). This is unlike linear correlation which depends on both. Rank correlation has been found to satisfy the first four properties (Embrechts, McNeil, and Strauman, 2002). Estimation of rank correlation requires the ordering of the sample for each variable. It does not require actual numerical values. The two most common rank correlation measures are Kendall's tau and Spearman's rho. These two rank correlation measures are both symmetric and take on values in the interval [-1, 1]. In addition, they are both invariant under strictly increasing transformations. Kendall's tau is the rank correlation measure used in this research.

Kendall's tau measures the concordance between two random variables. Concordance is best illustrated by an example. Suppose there are two points, the first point denoted by (x, y) and the second point denoted by $(\tilde{x}, \tilde{y})$. These two points are considered to be concordant if $(x - \tilde{x})(y - \tilde{y}) > 0$ and discordant if $(x - \tilde{x})(y - \tilde{y}) < 0$. This relationship is

$$\rho_\tau(x, y) = P\big((x - \tilde{x})(y - \tilde{y}) > 0\big) - P\big((x - \tilde{x})(y - \tilde{y}) < 0\big), \qquad (2.2)$$

where *P* represents the probability of either concordance $\big((x - \tilde{x})(y - \tilde{y}) > 0\big)$ or discordance $\big((x - \tilde{x})(y - \tilde{y}) < 0\big)$. The left hand side of equation (2.2) represents the scalar measurement for Kendall's tau. In other words, Kendall's tau is equal to the probability of concordance minus the probability of discordance.

2.4.3 Copulas

The concept of copulas is being used to overcome the shortcomings of linear correlation and to efficiently estimate rank correlation. Copulas model multivariate distributions. An extensive treatment of copulas can be found in (Accioly, and Chiyoshi, 2004; Ane, and Kharoubi, 2003; Rank, 2000; Schmidt, 2006; Trivedi, and Zimmer, 2005). Here, only a basic treatment of copulas is provided to lay the foundation for comparing the different measures in the context of geographical diversification. The origin of copulas can be traced back to the Sklar theorem (Sklar, 1959). The Sklar theorem shows that any joint distribution function may be decomposed into its *n* marginal distributions and consequently a copula. The copula is then considered to describe the dependence that exists between the variables (Patton, 2002).

Copulas conveniently allow for the separation of the marginal distributions from the joint distribution. This separation allows the flexibility of modeling each individual marginal distribution. Then the copula function is used to join the marginal distributions to obtain the multivariate distribution. Copula parameters are estimated through a Maximum Likelihood Estimation method of the form of

$$\hat{\delta}_2 = argmax_{\hat{\delta}_2} \sum_{t=1}^{T} \ln c\left(\hat{G}_x(x_t), \hat{H}_y(y_t), \delta_2\right) \quad (2.3)$$

where $\hat{\delta}_2$ is the estimated copula parameter, $argmax$ is the mathematical function that provides the argument associated with the maximum, ln is the natural logarithm, and $\hat{G}_x(x_t), \hat{H}_y(y_t)$ are the estimated marginal distributions for *x* and *y*. The copula parameter $(\hat{\delta}_2)$ provides the information necessary to estimate the copulas and tail dependence between the crop yields for the individual counties.

Of the different copulas that have been developed, four different copulas will be compared: Gaussian; Gumbel; Frank; and Clayton. The Gaussian copula is symmetric, whereas, the Clayton, Gumbel, and Frank copulas are asymmetric with additional probability in the tails of the distribution. The functional form for each of the copulas is given in Table A.3. The table illustrates the important characteristics of each one of the copulas used in this analysis. Copulas also provide an alternative method for estimating rank correlation. Kendall's tau can be defined by using the estimated copula. Using copulas Kendall's tau is defined as:

$$\rho_\tau(x,y) = 4\int_0^1\int_0^1 C(u_x,u_y)d(u_x,u_y) - 1. \tag{2.4}$$

The integral above can interpreted as the expected value of the function $C(u_x,u_y)$, where $u_x,u_y \sim u(0,1)$ are the uniform transformed random variables *x* and *y* and *C* is the joint distribution function for u_x and u_y.

2.4.4 Tail Dependence

Tail dependence is another alternative to analyze the relationship between two random variables. The concept of tail dependence is concerned with the magnitude of dependence that exists in the upper right quadrant or lower left quadrant tail of a bivariate distribution (Charpentier, and Segers, 2007). This concept is essential when studying the measure of dependence between extreme values (Cherubini, Luciano, and Vecchiato, 2004). A formal definition of an upper tail dependence coefficient is

$$\lambda_U = \lim_{t\to 0^+}\{G(x) > t|H(y) > t\}. \tag{2.5}$$

The estimate of λ_U is the upper tail coefficient. *G(x)* and *H(y)* represent the marginal distributions for *x* and *y*. The variable *t* represents the upper threshold limit. Similarly the lower tail coefficient can be defined by

$$\lambda_L = \lim_{t \to 0^+} \{G(x) \leq t | H(y) \leq t\}. \quad (2.6)$$

In this case, the variable *t* represents the lower limit threshold. These coefficients are interpreted as the probability that one margin exceeds a given high (low) threshold under the specified condition that the other margin exceeds the high (low) threshold. Or in other words, the coefficients examine how closely the tails of the data move together.

Copulas have been found to be useful when modeling tail dependence (Kallenberg, 2008). The tail dependencies of different parametric copulas emphasize different portions of the probability distribution. For this reason, the copulas used are the ones that capture tail dependence (McNeil, Frey, and Embrechts, 2005). The Gaussian copula is the only copula with zero tail dependence. The Student-t copula has symmetric tail dependence. Clayton, Rotated Gumbel, and Gumbel copulas have asymmetric tail dependencies. This research is concerned with the lower tails of the data and so the Clayton and Rotated Gumbel will be used to estimate the tail dependence coefficient. It has been shown that there is a simple method to calculate the tail dependence from the estimated copula for both the Clayton and the Rotated Gumbel copulas. For the Rotated Gumbel copula, the tail dependence is calculated by

$$\lambda_L = 2 - 2^{1/\delta}, \quad (2.7)$$

where λ_L represents the lower tail coefficient and δ represents the estimated copula parameter. The tail dependence for the Clayton copula is calculated by

$$\lambda_L = 2^{-1/\delta}, \tag{2.8}$$

where λ_L represents the lower tail coefficient and δ represents the estimated copula. The results of both measures are presented.

2.5 Yield Dependence Model Specification

To address the objectives, yield dependence functions of the following general form are estimated

$$d_{ijc} = f\left(Ele_{ijc}, Prec_{ijc}, Temp_{ijc}\right) + \varepsilon_{ijc} \tag{2.9}$$

where d_{ijc} is the county level yield dependence coefficient between county *i* and county *j* for crop combination c, ε_{ijc} is the error term, and the remaining variables are differences in absolute value between the two counties *i* and *j* in elevation (*Ele*) in feet, annual precipitation (*Prec*) in inches, and annual temperature (*Temp*) in Fahrenheit. Dependence functions are estimated individually for wheat, cotton, and sorghum.

Three specifications of equation (2.9) are estimated, linear, log-linear, and quadratic. A log-linear instead of a log-log form is estimated because some of the correlations are negative. Statistical tests are conducted to determine which functional form provides the "best" fit. Summary statistics for the variables used in the regression equation are found in Table A.3.

Before calculating county yield dependence, the data is detrended to eliminate time trend components. Yield data are detrended using a simple linear trend model (Table A.1)

$$Y_{ntc} = \alpha_{ntc} + \beta_{ntc}t + \varepsilon_{ntc} \tag{2.10}$$

where Y_{ntc} is county yield from county n in year t for crop c, α and β are coefficients to be estimated, t represents the year with t = 1, 2, …., T, and ε_{ntc} is the error term. The significance of the coefficient β_{ntc} is used to determine whether a trend is present. Approximately 50% of the counties show a significant trend in the wheat yield data (Table A.1). Cotton county yields exhibited a trend in only 10% of the counties, whereas, 60% of the sorghum counties exhibited a significant trend (Table A.1). To be consistent, all yields are detrended. Detrended county level yields (the residuals from the trend equation) are used to calculate the linear correlation and to estimate the copulas.

2.5.1 Elasticities

To illustrate how changes in the independent variables affect yield correlations, point elasticities are calculated as follows

$$\varepsilon_{xc} = \left|\frac{\partial d_{ijc}}{\partial x}\frac{\bar{x}}{\hat{d}_{ijc}}\right| \tag{2.11}$$

where ε_{xc} is the elasticity for independent variable of interested (elevation, temperature, or precipitation) x and crop combination c, d_{ijc} represents the estimated dependence function (equation 2.9), ∂ represents the partial derivative, and $\hat{d}_{ijc}$ is the estimated correlation function evaluated at the mean of the independent variables. Elasticities give the percentage change in yield correlations for a given change in the independent variables. Because the functions are not constant elasticity functions, elasticities vary based on the value of the independent variables. Following standard procedures, the elasticities are calculated at the mean of the independent variables. Because elasticities are normally defined for when both variables dijc and x are positive, but this is not always the case with the dependence measure, the absolute value of the elasticities is presented. The elasticities are not used to provide direction just magnitude of the relationships.

2.6 Linear Correlation Results

Using the detrended yields and the geographical data for each county, equation (2.9) was estimated for each of three commodities. The quadratic functional form provided the best fit for each of the three model selection criterion, Akaike Information Criterion, Bayesian Information Criterion, and adjusted R^2 (Table A.4). For space considerations and ease of discussion, estimation coefficients are discussed only for the quadratic functional form (the coefficients of the linear and log-linear regressions can be found in Tables A.5 and A.6). The results of the individual analysis using Pearson correlation will be followed by a discussion of both the copula estimated dependence and the tail dependence results.

2.6.1 Wheat

Using the quadratic specification, all wheat variables are significant (at the 10% level or less). As expected, all coefficients of the linear terms are negative (Table A.7). The coefficient of the squared term for temperature is positive, whereas, the coefficients for the squared terms of elevation and precipitation are negative. These results provide support for the hypothesis that there is generally an inverse relationship between yield correlations and geographic variables in the relevant range of the independent variables. In other words, correlation between yields is reduced as absolute differences in both spatial and climate variables increases. The negative squared term combined with the negative linear term is indicative of a concave shape (correlations decreasing at a decreasing rate), while the positive squared term combined with the negative linear term is indicative of a convex shape (correlations decreasing at an increasing rate).

Absolute value of the elasticities associated with percentage changes in yield correlations for each of the geographical variables are calculated at their means to an indication of the relative effect of each variable (Table A.8). The elasticity of temperature and precipitation are 1.28 and 1.09. These elasticities can be interpreted as a 1% change in either variable leads to a 1.28% and 1.09% change in the correlation between county wheat yields. The estimated elasticity for elevation is 0.18; a 1% change in elevation leads to a 0.18% change in correlation. These elasticities suggest temperature and precipitation have a greater effect on yield correlations than changing elevation.

2.6.2 Cotton

All variables are significant at the 10% level for except for the squared elevation term and the elevation/precipitation interaction term (Table A.7). The linear terms for temperature and precipitation are negative, while the linear term for elevation is positive. The squared terms are the opposite, temperature and precipitations are positive while elevation is negative. The positive squared term indicates a convex shape for temperature and precipitation, whereas, the negative squared term combined with the positive linear term indicates a concave shape for elevation.

Estimated elasticities for temperature and precipitation are 0.19 and 1.23 (Table A.8). The estimated elasticity for elevation is 0.76. Dry land cotton is more sensitive to changes in precipitation than changes in temperature or elevation. The elevation elasticity estimate must be used with caution because of the lack of significance of two of the elevation terms. These elasticity estimates are consistent with the idea; precipitation plays a larger role in decreasing cotton yield correlations than elevation.

2.6.3 Sorghum

Elevation and temperature linear terms are significant at the 10% level, while the linear precipitation term is not significant (Table A.7). The squared term for temperature is the only significant squared variable. The elevation/precipitation interaction and the precipitation/temperature interaction terms are insignificant. These estimates suggest that changes in precipitation levels have little effect on sorghum yield correlations but changes in temperatures are more relevant.

Elevation and temperature have estimated elasticities of 0.42 and 1.40 (Table A.8). Precipitation has an elasticity of 0.06. The precipitation elasticity must be viewed with caution, because of the lack of statistical significance of the coefficients. Lack of statistical significance can partially be explained by the concentration of sorghum production and by sorghum’s ability to handle lower precipitation levels. These elasticity estimates illustrate that changes in temperature have the greatest impact on yield correlations. For every one degree change in temperature, the yield correlations will change by 1.4%.

2.6.4 Three Crop Comparisons

A comparison of all three commodity elasticities illustrates some interesting characteristics of the different crops. Cotton is the most sensitive to elevation changes with an estimated elasticity of 0.76. Wheat is the least sensitive to elevation changes with an estimated elasticity of 0.18. Wheat and sorghum are both sensitive to changes in temperature. Their elasticities associated with temperature are both greater than one (1.28 for wheat and 1.40 for sorghum). Cotton is the most sensitive to changes in precipitation with an elasticity of 1.23. Wheat is also sensitive to changes in precipitation with an elasticity of 1.09. Sorghum was affected little by changes in precipitation levels with an elasticity of 0.06.

2.7 Copula Results

The estimation of the copulas was done for all three crops. The copulas estimated were: Gaussian, Frank, Clayton, and Gumbel copulas are estimated for all three crops. The estimated copula parameters are used to calculate the corresponding Kendall's tau. The Kendall's tau measure is then used as the dependence measure in the model specification. To maintain consistency with the previous results, the quadratic specifications are presented. Three criteria will be used to compare the estimation results using the different dependence measures: statistical significance, sign, and magnitude.

2.7.1 Wheat

The statistical significance and signs of the coefficients for wheat yield dependence do not change with the different dependence measures (Table A.9). All geographical and climate independent variables are significant (at the 10% level or less) except for the squared elevation term. As expected, the coefficients of the linear terms are negative. The coefficient of the squared term for temperature is positive, whereas, the coefficients for the squared terms of elevation and precipitation are negative. These results provide support for the hypothesis that there is generally an inverse relationship between yield correlations and differences in geographic and climate conditions In other words, dependence between yields is reduced as absolute differences in both geographical and climate variables increases. The negative squared term combined gives a concave shape (correlations decreasing at a decreasing rate), while the positive squared term provides a convex shape (correlations decreasing at an increasing rate).

Given that the sign and statistical significance of the independent variables do not changed among the alternative dependence measures, the change in magnitude is another criterion for comparing the measures. To compare the magnitude, the absolute value of the elasticities, evaluated at the mean of all the variables are used (Table A.10) along with estimated dependencies at different levels of the independent variables (Figure A.2). Magnitudes of the elasticities vary among the different measures with linear correlation and Gaussian copula elasticities generally being larger than for the other copula measures Elevation elasticity varies only slightly across the different measures. The largest difference in elasticities is between linear correlation (0.18) and the Gumbel Copula (0.07).

In Figure A.2, the dependence functions are graphed with two of the three independent variables are held constant at their mean with the remaining variable varying from its minimum to maximum within the data. Because of differences in the estimated dependencies, different scaling is used in the different graphs in Figure A.2. As such, caution is advised in comparing between the graphs. With precipitation and temperature held constant at their means, the dependence functions as elevation changes are similar for all measures. The linear correlation however decreases slightly faster than the other measures. The reason for this is because the magnitude of the squared elevation term is the greatest for linear correlation and consequently it is decreasing faster than with the other measures.

Temperature elasticities have a larger difference in magnitude (Table A.10). The linear correlation (1.28) and Gaussian copula (1.33) elasticities are similar, but they differ from the elasticities associated with the Frank (0.77), Clayton (0.69), and Gumbel (0.70) copulas. The different estimated dependence functions as a function of temperature are shown in Figure A.2. All five dependence measures exhibit the same shape. The Gaussian copula and linear correlation are similar in magnitude, whereas, the graphs of the remaining copulas are similar in magnitude at the beginning but show some divergence from each other as the change in temperature becomes larger.

Mean precipitation elasticities shows relationship similar to that of the temperature elasticities. The linear correlation and Gaussian have the same elasticity (1.09), whereas, the Frank (0.56), Clayton (0.51), and Gumbel (0.50) copulas all have estimated elasticities that are roughly half the magnitude of the linear and Gaussian elasticity. The five measures illustrate the same concave shape (dependency decreasing at a decreasing rate). Graphs of the dependence measures with respect to precipitation reflect the similarity of Gaussian and linear correlation functions. The Gaussian and linear correlation functions are decreasing faster than the remaining three copula measures. The remaining functions are also similar to each other over the range of the data.

For all estimated functions, the magnitudes of the elasticities have the following order. Elevation elasticities are the smallest and temperature elasticities are the largest. This ordering indicates wheat dependencies are most sensitive to temperature differences and least sensitive to elevation changes. Sensitivity of dependencies to precipitation differences falls in between elevation and temperature.

2.7.2 Sorghum

Unlike wheat, there are some differences in the sign and statistical significance of the geographical and climate variables in the sorghum estimations (Table A.11). The linear precipitation coefficient is negative for linear correlation, Clayton, and Gumbel copulas, whereas it is positive for the Gaussian and Frank copulas. The squared precipitation term is negative for all dependence specifications except the Clayton copula. The squared elevation term in insignificant under the linear correlation specification, but is significant for all four copula measures. The elevation/precipitation interaction term is insignificant in the linear correlation and Gaussian copula functions, but significant for the remaining three copulas. The signs of the coefficients for the remaining variables are the same for all specifications.

Also differing from the wheat estimations, the linear correlation and Gaussian copulas do not provide similar elasticity magnitudes (Table A.10). The elevation elasticity is the greatest using the Gaussian copula (3.17), whereas, the linear correlation (0.61) has the lowest elasticity. The other three copulas have estimated mean elevation elasticities near 1. For changes in elevation, the linear correlation and Gaussian estimated dependence functions different not only from each other but also from the other three estimated functions (Figure A.3). Although the five dependence functions have the same basic shape, the linear correlation function decreases at a slower rate than the other four dependence functions. The linear correlation function also begins to increase for larger differences in elevation, whereas the other measures show decreasing dependence over the entire range of the data.

Temperature mean elasticities exhibit a relationship similar as to that shown with elevation with the Gaussian elasticity being the largest and the linear being the smallest (Table A.10). The elasticities associated with the copulas except for the Gaussian copula (3.47) are close to one. The temperature elasticity for linear correlation is also near one. The functions graphed with respect to temperature are similar to the functions with respect to elevation in terms the order of the functions (Figure A.3). The Gaussian copula and linear correlation dependence functions have steeper slopes than the other three copulas. The Frank, Clayton, and Gumbel all remain similar over the range of the data.

Precipitation mean elasticities and graphs must be used with caution, but are presented here for completeness. In none of the estimations are the linear and the squared precipitation terms statistically significant. This lack of significance may help explain the intuitively unpleasant graphed dependence functions over the range of precipitation differences (Figure A.3). All five copulas exhibit an increasing dependency that is increasing at a decreasing rate. Linear correlation, on the other hand, exhibits decreasing dependency at a decreasing rate. The lack of significance of the precipitation variables may be because of the relative drought tolerance of sorghum.

Unlike wheat, the magnitudes of the mean elasticities do not have a consistent ordering. Precipitation mean elasticities are the smallest for all dependence measures, again possibly because of sorghums drought tolerance. Elevation and temperature mean elasticities' order depends on the measure.

2.7.3 Cotton

Similar to sorghum and unlike wheat, there are some differences in the signs and statistical significance of the estimated coefficients associated with cotton yield dependencies (Table A.12). The squared temperature and precipitation terms are significant in the linear correlation model but statistically insignificant in the copula models. Precipitation/temperature interaction term is significant in the Clayton copula model but insignificant for the remaining specifications. Signs of the coefficients vary among the models for the precipitation squared and precipitation/temperature interaction variables.

The linear correlation mean elasticities for elevation, temperature, and precipitation are larger than any of the copula based mean elasticities (Table A.10). Mean elasticity associated with the Gaussian copula is the second largest for all the variables. For elevation, the copula based elasticities are approximately one-half (between 0.4 and 0.6) of the linear correlation mean elasticity. The estimated dependence functions with respect to elevation are similar in shape (Figure A.4). All the functions are increasing at a decreasing rate but start out negative. The linear correlation function has a slightly steeper slope than the copula measures.

Mean temperature elasticities are smaller for temperature and the copula functions relative to linear correlation than for elevation. Copula elasticities are approximately 0.3 to 0.6 of the linear correlation elasticity. The dependence functions associated with changes in temperature exhibit an opposite relationship as seen with changes in elevation (Figure A.4). All five dependence functions are decreasing at an increasing rate. The linear correlation and Gaussian functions are similar, but the linear correlation function decreases more rapidly than the Gaussian function. The remaining three copula functions are all similar to each other.

Precipitation mean elasticities follow a pattern similar to that seen in the elevation mean elasticities for cotton among the different specifications. The copula elasticities for precipitation are approximately 0.4 to 0.6 of the linear correlation elasticity. Changes in precipitation exhibited similar changes in the dependence functions (Figure A.4). The main difference can be seen in the Clayton copula. All the functions are decreasing functions with respect to precipitation, but the Clayton copula function has a negative squared precipitation term. Whereas the other functions all have a positive squared precipitation terms. The results of this shape difference can be seen as the precipitation changes become large.

Cotton mean elasticities show a distinct ordering between the geographical and climate variables among the different dependence measures. Temperature has the highest mean elasticity and precipitation the smallest elasticities.

2.7.4 Tail Dependence

The tail dependence coefficient is estimated using the results of the copula estimations. Consistent with the previous estimations, only the results of the quadratic estimation are presented. Because the tail dependence coefficient is constrained on the range 0 to 1, a Tobit model will be used to estimate the equation. In addition, because the tail dependence range is different than the dependence measures used above, it is difficult to compare directly to the other dependence measures. The Clayton and Gumbel copulas are used to estimate the tail dependence for the three different crops. The tail dependence in wheat is consistent in sign and statistical significance to the other estimations (Table A.9). The magnitudes of the elasticities are also very similar to the copula dependence measures (Table A.10). The results of the sorghum tail dependence are consistent in both sign and significance with the copula dependence estimations except for the squared elevation term (Table A.11). Using Gumbel tail dependence, the coefficient associated with the squared elevation term is insignificant while it was significant under the other copula estimations. The magnitude of the elasticity is similar to the other copula based estimations (Table A.10) being around 1 for elevation and temperature and 0.15 for precipitation. The cotton tail dependence estimations are also consistent with the copula estimations (Table A.12). The magnitudes of the tail dependence measures are once again similar to the other copula dependence measures

(Table A.10). The elevation elasticities are around 0.80, 1.30 for temperature, and 0.55 for precipitation.

2.8 Conclusions

Although the issue of geographical diversification has not been extensively studied, it provides an opportunity to examine interesting risk management issues. This study provides a starting point for analyzing the potential impact on yield dependencies from geographical diversification. This study illustrates the expected results that yield dependencies vary by geographical and climate changes. Quantification of the relationships between yield correlations and geographical and climate variables allows the next step of geographical diversification to be undertaken, namely examining how geographical diversification will impact risk and profitability of agricultural enterprises. Elasticity estimates suggest on a percentage basis, changes in temperature and precipitation have the greatest effect on dependencies.

This research has the secondary objective of analyzing the impact of using alternative dependence measures. Four parametric copulas, Gaussian, Frank, Clayton, and Gumbel, are compared to the more typical measure of dependence, Pearson's linear correlation. The effect of geographical and climate variables differed between the different dependency measures among the three different crops. Implementing alternative dependency measures changed the statistical significance and the signs of the coefficients in the sorghum and cotton dependence functions. Copula based elasticities are consistently less than the linear correlation elasticities for wheat and cotton. For sorghum, however, the copula based elasticities are generally larger.

These differences have potential implication for applied studies. Given the differences in the effect of geographical and climate variables on yield dependencies, a logical question is how do this differences relate to changes in “optimal” risk management strategies, if any. Results from the different dependency functions indicate there will be differences. Only applied farm level modeling on geographical diversification will provide a quantitative answer. Inferences, however, strongly suggest one should not take the estimation of yield dependencies lightly. Applied studies using real world data in which the underlying distributions are unknown, may consider using several different dependencies. Another implication why are there differences among the crops beyond obvious crop characteristics. Wheat yield dependencies are more robust than the other crops. One must ask is this a function of the crop itself or the production locations. The wheat data set is more robust in terms of the number and variability in locations grown. This is an area future research should examine.

Satisfying the objectives of this research establishes a foundation for both researchers and producers to better understand the impacts of alternative dependency measures and geographical diversification. An extension of this research is to examine non-parametric copulas. May be the most important extension is to examine the robustness of the different dependence measures for various underlying distributions in a Monte Carlo framework.

3. INCORPORATING ASYMMETRIC DEPENDENCE AND RISK MEASURES INTO PORTFOLIO OPTIMIZATION: AN APPLICATION TO GEOGRAPHICAL DIVERSIFICATION

3.1 Introduction

Portfolio theory has provided individuals with a means to measure and manage risk. The risk measure used in traditional portfolio problems has been variance (Markowitz, 1952). Diversification provides a method to manage risk based on minimizing variance and covariance between assets. Variance is a valid risk measure and linear correlation is the appropriate measure of dependence when the returns of the assets in the portfolio are normally distributed (Szegö, 2005). The existence of normally distributed asset returns in real world situations has been shown to be limited (Just, and Weninger, 1999; Sun, et al., 2009). Alternative risk measures and dependency measures have been developed to account for non-normal data (Stoica, 2006).

When variance is used as the risk measure, all risk is treated the same. Upside risk is penalized the same as downside risk. This symmetric view of uncertainty is often considered to be counter intuitive to real-world situations (Alexander, and Baptista, 2004). Upside risk is often considered to be the riskless opportunities for unexpectedly high returns. Individuals are often not concerned with the upside risk but with the downside risk where downside risk is measured as the volatility below the individual's target return.

The use of downside risk measures in portfolio settings has been embraced by the corporate finance and banking industry (Acerbi, 2007; Alexander, and Baptista, 2002; Artzner, et al., 1999; Buch, and Dorfleitner, 2008). For example, the Basel Committee on Banking Supervision utilizes a downside risk measurement in their evaluation of capital standards for banks (BIS, 2004). They use the downside risk measure estimated by value at risk (VaR). Downside risk measures such as VaR have only been used in a number of agricultural applications (Manfredo, and Leuthold, 2001). The use of VaR as a downside risk measure has also been found to be problematic (Artzner, et al., 1999). VaR is subject to many of the same limitations as variance. It is only valid when the returns in a portfolio are normally distributed (Artzner, et al., 1999).

Recent developments in additional downside risk measures have led to a class of risk measures referred to as coherent risk measures (Rockafellar, and Uryasev, 2000). A risk measure is considered to be coherent if it satisfies the properties of translation invariance, subadditivity, positive homogeneity, and monotonicity (Acerbi, 2007). One risk measure that has been found to satisfy these properties is conditional value at risk (CVaR). CVaR measures the expected value of losses for a given probability conditional on losses less than or equal to VaR for that probability.

The two most popular methods of computing downside risk measures are the variance-covariance method and Monte Carlo simulation (Duffie, and Pan, 1997). The variance-covariance method is subject to the assumption of multivariate normality. The Monte Carlo method provides more flexibility and allows the incorporation of copulas to model the multivariate relationships.

Copulas provide more flexibility in modeling the dependence between net returns generated at different location and thus overall risk faced by a producer. Copulas allow the modeling of the marginal distributions separately from the multivariate distribution. This provides the flexibility to fit each marginal distribution to the most appropriate distribution. Once the marginal distributions have been specified, the copula function is fit to the multivariate distribution. Or in other words, the copula couples together the marginal distributions to the multivariate distribution. The copula function can capture non-linear dependence and provide a more accurate picture of the relationship between assets.

The development of alternative risk and dependence measures, such as CVaR and copulas, has opened the door to reexamine the traditional portfolio problem. In agricultural settings, the applications alternative risk measures is particularly applicable because of the changing nature of the agricultural industry. In particular, traditional family farms have been replaced by commercial agricultural enterprises that are focusing more on specialized production than crop mix diversification (Mishra, El-Osta, and Sandretto, 2004). This transition from the traditional family farm to an industrialized production enterprise has opened new ways for risk diversification in agribusinesses (Boehlje, and Lins, 1998; Vedenov, and Barnett, 2004; Zhu, Ghosh, and Goodwin, 2008). To help manage the risk, some of these commercial enterprises have begun to diversify geographically (Larsen, 2008).

Previous analysis of geographical diversification research has provided no clear answers as to its effectiveness as a risk management tool (Davis, et al., 1997; Krueger, Salin, and Gray, 2002; Nartea, and Barry, 1994). However, part of the confusion may lie in weaknesses of the methodology used in the analysis. Traditional reliance on variance minimization or mean-variance optimization criterion may lead to diversification strategies that penalize upward deviations from the target revenue. In addition, the dependence between revenue streams generated at different locations is typically measured through linear correlation, which may affect the results when the joint distribution of returns is non-elliptic (e.g. not a multivariate normal). The objective of this study is to analyze the portfolio allocations when these alternative dependence and risk measures are used in a single period portfolio problem. This is done by analyzing the portfolio choice using the CVaR/copula method versus CVaR/multivariate normal assumption. The efficient set of both portfolio models are presented as well as comparisons of portfolio allocations for different levels of risk aversion.

The results from both models provide a means to analyze the effectiveness of geographical diversification as a way to manage risk. More specifically, annual net returns from dry land wheat production are collected for three regions – Texas, Colorado, and Montana. A copula-approach is used to calculate the joint distribution of the returns at each location and thus the combined return of the enterprise. This study advances previous research in two important ways. First, it combines copulas (the dependency structure model) and CVaR (risk management criterion) to analyze risk management problems in agriculture. Furthermore, geographical diversification between states is addressed.

3.2 Geographical Diversification as a Risk Management Strategy

Previous studies of geographical diversification in agriculture have produced somewhat contradictory results. Nartea and Barry (1994) analyzed the costs and returns of geographical diversification in Central Illinois to determine whether geographical diversification was a legitimate risk management strategy for individual grain growers. After comparing the increases in revenues received with increases in transportation and monitoring costs and losses due to poor machinery coordination, the authors concluded that there was no realizable gain from diversifying geographically in Central Illinois.

Davis et al.(1997) examined the impact of geographical diversification on peach orchards in Georgia. The authors argued that weather related production risks could be reduced due to spatial scattering of production activities. Furthermore, peach production provided a unique example because of the lack of alternative risk reduction instruments such as government support programs and financial instruments. Using a stochastic production function, the authors determined the variability of yield that could be reduced by geographically scattering peach orchards. They found that for every mile increase in distance between orchards, correlation between yields dropped by 2%. The authors concluded that implementing geographical diversification was a legitimate risk reduction strategy and geographical diversification could also enhance the long-term sustainability of the peach production.

3.3 Risk Measures and Copulas

Resource allocation problems traditionally utilize portfolio methods in order to determine how to best diversify resources (Crisostomo, and Featherstone, 1990;

Harwood, et al., 1999; Hennings, Sherrick, and Barry, 2005). The portfolio analysis, in turn, relies on the mean-variance optimization framework introduced by Markowitz more than 50 years ago (Markowitz, 1952). Since then numerous studies highlighted shortcomings of the mean-variance optimization and suggested alternative decision criteria (Alexander, and Baptista, 2004).

3.3.1 Coherent Risk Measures

The finance industry has embraced and utilized the value-at-risk (VaR) criterion (Jorion, 1996), however the applications of VaR in agricultural economics has been limited (Manfredo, and Leuthold, 2001). VaR is a convenient way to assess the magnitude of a variable of interest associated with a given probability of occurring. For instance, if applied to farm revenues, VaR at 5% indicates the level below which the revenues would drop less than 5% of the time (i.e. revenues will be greater than that 95% of the time). The advantages of VaR are its simplicity and intuitive interpretation. Recent research however, has shown that VaR does not possess the properties of a coherent risk measure, because it does not satisfy sub-additivity condition (Acerbi, 2007; Artzner, et al., 1999). A particularly troubling implication for portfolio optimization is that the VaR of a portfolio of two securities may be greater than the VaR of each individual security (Alexander, and Baptista, 2004). Furthermore, VaR is also shown to lead to erroneous results when the data is not normally distributed (Stoica, 2006). The effectiveness of risk management relies on the effectiveness of the risk measures involved. A current trend in risk management is that of using coherent risk measures.

Artzner (1999) proposed some properties that risk measures should have. These have become known as the coherent risk measures. The properties of the risk measure ϕ are:

1. Translation Invariance: In words, this means that by adding or subtracting a deterministic quantity l to a position leading to a loss L the capital requirements are altered by exactly that quantity. Mathematically:

   $\phi(L+l) = \phi(L) + l$

2. Subadditivity: This is essential to further the argument that risk can be reduced by diversification. For example, consider the losses from two different revenue streams, L_1 and L_2. The combined risk of the two revenue streams $\left(\emptyset(L_1 + L_2)\right)$ is less than or equal to the sum of the risks associated with each individual revenue stream $\left(\emptyset(L_1) + \emptyset(L_2)\right)$. Mathematically:

   $\emptyset(L_1 + L_2) \leq \emptyset(L_1) + \emptyset(L_2)$

3. Positive Homogeneity: This property is key because combined with property 2 (subadditivity), the risk measure ϕ is indeed convex. Mathematically:

   $\phi(nL) = \phi(L + \ldots + L) \leq n\phi(L)$

4. Monotonicity: This property can be stated as positions that lead to higher losses in every state of the world require more risk capital.

If a risk measure satisfies these four risk properties, it is considered to be coherent.

An alternative to VaR that has been gaining popularity and has been shown to be coherent is the Conditional Value-at-Risk (CVaR). CVaR measures the expected value of losses for a given probability conditional on losses less than or equal to VaR for that probability. The CVaR has been found to be a more consistent measure of risk than VaR and generally resulted in more efficient portfolio choices (Alexander, and Baptista, 2004; Rockafellar, and Uryasev, 2000). CVaR, therefore, is used as a measure of risk reduction due to geographic diversification for the purposes of this study. CVaR can be defined as the expected loss given that the loss is greater than or equal to the VaR. The CVaR can be formally defined as

$$CVaR_{\alpha}(Y) = E[Y|Y \geq VaR_{\alpha}(Y)]. \tag{3.1}$$

The variable (Y) in equation (3.1) can be considered to be the random expected loss. The fixed level α is used to define the α-quantile for VaR.

3.3.2 Copulas

Calculation of CVaR requires knowledge of the cumulative distribution function of portfolio returns, which in turn depends on the joint distribution of returns of all assets included in the portfolio. Traditional approach to this type of problems relied heavily on the multivariate normal distribution (Markowitz, 1952). However, the assumption of normality for agricultural prices and yields has been shown to be inconsistent (Goodwin, and Ker, 2002; Just, and Weninger, 1999). Copulas are an alternative method of modeling joint distributions that has been gaining popularity in financial literature including portfolio analysis (Alexander, Baptista, and Yan, 2007; Alexander, Coleman, and Li, 2006; Bai, and Sun, 2007; Bouyé, et al., 2001; Clemen, and Reilly, 1999; Dias, 2004; Hennessy, and Lapan, 2002). The main advantage of copulas is their flexibility in specifying the marginal distributions of prices and yields while properly specifying the dependency that exists between them. While copulas have been used in finance for quite some time the applications of copulas in the agricultural literature are recent (Vedenov, 2008; Zhu, Ghosh, and Goodwin, 2008). This study uses copula methodology to model the joint distribution of random variables of interest, which is needed for calculation of CVaR criteria.

The application of copulas to modeling multivariate distributions is described in detail in numerous books and research articles (Cherubini, Luciano, and Vecchiato, 2004; Nelsen, 2006). For the purpose of this paper, a basic treatment of copulas is included to the extent required by the analysis[1]. A copula function is formally defined as an n-dimensional multivariate cumulative distribution function defined on the n-dimensional unit cube $[0,1]^n$ with the properties (i) $C(u_1,\ldots,u_n)=0$ if any $u_i=0, i=1,\ldots,n,$ and (ii) $C(1,\ldots,1,u_i,1,\ldots,1)=u_i$ for any $u_i, i=1,\ldots,n.$ The copulas are related to joint distributions through Sklar's theorem, which (in two-variable case) postulates that if H is a joint distribution function with margins F and G, then there exists a copula function C such that for all $x, y \in \overline{R}$, $H(x,y)=C(F(x),G(y))$ (e.g. (Nelsen, 2006).

The Sklar theorem allows one to construct joint distribution of several random variables based on their marginal distributions and a copula. By definition there are an infinite number of copula functions and thus an infinite number of joint distributions that may be generated for given marginals. Various copula families have been used in risk research (e.g. Gaussian, Archimedean, etc. (Hennessy, and Lapan, 2002)). Three

[1] For a complete review of copula theory refer to Joe, H. 1997. *Multivariate Models and Dependence Concepts.* D. R. Cox, ed. London, UK: Chapman & Hall, Nelsen, R. B. 2006. *An Introduction to Copulas*. 2nd Edition. Springer Series in Statistics. New York: Springer.

copulas from Archimedean family (Clayton, Frank, and Gumbel), Gaussian, and T copula are used in this research.

3.3.2.1 Gaussian Copula

The Gaussian Copula is an extension of the multivariate normal distribution but it can be used to model multivariate data that may exhibit non-normal dependencies and fat tails. The Gaussian Copula is formally defined as

$$C(u_1, \ldots, u_n; \Sigma) = \Phi^K(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n); \Sigma), \tag{3.2}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and Σ is the variance-covariance matrix. In the two-dimensional case, the Gaussian copula density can be written as

$$c(u, v) = \frac{1}{\sqrt{1-\rho^2}} exp\left(\frac{\left(\Phi^{-1}(u)\right)^2 + \left(\Phi^{-1}(v)\right)^2}{2} + \frac{2\rho\Phi^{-1}(u)\Phi^{-1}(v) - \left(\Phi^{-1}(u)\right)^2 - \left(\Phi^{-1}(v)\right)^2}{2(1-\rho^2)}\right), \tag{3.3}$$

where ρ is the linear correlation between the two variables and $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution. One of the useful features of the Gaussian copula is that it is parameterized by a single parameter (correlation coefficient) which can be estimated from historical data.

3.3.2.2 T Copula

The t copula is derived from the multivariate standardized t-Student distribution. It can be defined as

$$C(u_1, \cdots, u_n; \textstyle\sum, \nu) = T_{\sum,\nu}\left(t_\nu^{-1}(\hat{u}_1), \cdots, t_\nu^{-1}(\hat{u}_n)\right)', \tag{3.4}$$

where $T_{\Sigma,\nu}$ is defined as the standardized multivariate Student's t distribution function, Σ is the correlation matrix, and ν are the degrees of freedom. $t_\nu^{-1}(\hat{u})$ is used to denote the inverse of the Student's t cdf function. In the two dimensional case, the T copula density can be written as

$$c(u,v) = |\Sigma|^{-1/2}\frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left[\frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\right]^n \frac{\left(1+\frac{\zeta'\Sigma^{-1}\zeta}{\nu}\right)^{-\frac{\nu+n}{2}}}{\prod_{i=1}^{n}\left(1+\frac{\zeta_i^2}{2}\right)^{-\frac{\nu+1}{2}}}, \tag{3.5}$$

where ζ is the vector of the T-student univariate inverse distribution functions. Both of these copulas are well formulated to take beyond the bivariate case.

3.3.2.3 Archimedean Copulas

Archimedean copulas are defined by

$$C(u,v) = \varphi^{-1}\big(\varphi(u) + \varphi(v)\big), \tag{3.6}$$

where φ is the generator of the copula. One of the most appealing features of Archimedean copulas is the relationship between the generator of the copula φ, and Kendall's tau. This relationship can be defined by

$$\tau = 1 + 4\int_0^1 \frac{\varphi_\alpha(t)}{\varphi_\alpha(t)}\,dt, \tag{3.7}$$

where τ is Kendall's tau. This provides a method of comparing rank correlation measures using different dependence structures. Three specific Archimedean copulas are used in this research, Clayton, Frank, and Gumbel Copulas. The Clayton copula is an asymmetric copula and exhibits greater dependence in the lower tail. The Frank copula on the other hand is a symmetric copula and weights the tails of the data equally.

The Gumbel copula is an asymmetric copula and exhibits greater dependence in the upper tail.

Traditionally, the implementation of copulas involves three steps: select and construct a copula, estimate the parameters associated with the copula, and sample from the parameterized copula. The Gaussian and t-copula are used in this research. The details on their construction were discussed in the previous section. Copula parameters are estimated through a maximum likelihood estimation method of the form of

$$\hat{\delta}_2 = argmax_{\hat{\delta}_2} \sum_{t=1}^{T} \ln c\left(\hat{G}_x(x_t), \hat{H}_y(y_t), \delta_2\right), \tag{3.8}$$

where $\hat{\delta}_2$ is the estimated copula parameter, argmax is the mathematical functions that provides the argument associated with the maximum, ln is the natural logarithm, and $\hat{G}_x(x_t), \hat{H}_y(y_t)$ are the estimated marginal distributions for *x* and *y*. To avoid any distributional assumptions, a non-parametric distribution is used for the marginal distributions. The final step is to draw random numbers from the estimated copula. Using this framework, a large sample of random values can be generated to be used as input into the optimization routine.

3.4 Methodology

The first step in developing the optimization model to evaluate the risk impacts of geographical diversification is to define the returns. This research is focused on the specialized production of dry land wheat in three separate production regions. In the initial step, it is assumed that prices (p) and yields (y) are deterministic. The multiplication of prices and yields results in the gross revenue of wheat production (GR) defined below in equation 3.9.

$$GR = w \cdot p \cdot y. \tag{3.9}$$

In the above case, w represents the amount of acreage devoted to production. In this single farm case, that is simply defined as the base acreage allocated to production (5,000 acres in this case). The next step is to incorporate the costs of production into the revenue equation to generate the profit.

$$\pi = GR - cw. \tag{3.10}$$

Profit (π) is defined as the gross revenue less the cost of production (c) that is a function of the number of acres in production (w). In the case that the producer may operate in more than one location, the following modifications are made to the profit equation.

$$\pi = \sum_{i=1}^{n} [w_i \, p_i y_i - c_i w_i], \tag{3.11}$$

where w_i represents the amount of acres allocated to each geographic location *i* with *n* possible locations. In this research, *n* represent the three possible growing regions. For simplicity sake, it is assumed that the acreage allotment is continuously divisible among locations. This allows the decision maker to allocate any given number of acres to any of the three regions given it satisfies the total acreage constraint. In addition $p_i y_i$ represent the prices and yields from each location and $c_i w_i$ is the cost equation that is associated with each geographic region.

The specification of equation (3.11) allows the formulation of an optimization problem that consists of maximizing profit subject to some constraints. Standard portfolio theory relies on the statistical measure of variance as the risk measure. This is modified by implementing CVaR as the risk measure. This allows the maximization of profit subject to reducing the "downside" risk. This has the obvious benefit of only penalizing losses and not "upside" risk, which is one of the drawbacks of the mean-variance optimization approach.

Formally, the portfolio optimization problem with the CVaR risk constraint can take two forms. The first consists of the CVaR function as the objective function. In this case, the CVaR is being minimized subject to a certain level of returns. Alternatively, the CVaR could be incorporated as a risk criterion and the objective function would consist of maximizing returns. This is the formulation that is used in this paper and takes the form of:

$$Max\ \pi = \sum_{i=1}^{n}[w_i p_i y_i - c_i w_i]. \tag{3.12}$$

Subject to

$$CVaR(w_i; \alpha) \leq \pi_0 \tag{3.13}$$

$$\boldsymbol{\sum_{i=1}^{n} w_i \leq A} \tag{3.14}$$

where π in equation (3.12) represents the sum of net returns from each geographic location i and w_i is the acreage allocation to each geographic location. Equation (3.13) represents the CVaR constraint which is a function of the acreage allocation for each region (w_i) and a confidence level α. π_0 is the minimum level of returns specified in the model. In the case of geographic diversification, this minimum level of returns could be calibrated to account for the additional costs associated with producing in multiple regions. Equation (3.14) represents the land constraint. The total amount of acreage $\sum_{i=1}^{n} w_i$ has to be less than or equal to the land endowment *A*.

3.5 Data

To analyze the effectiveness of geographical diversification as a risk management strategy, three geographically distinct areas were chosen based on harvesting windows and distance criteria — Pampa, TX, Akron, CO, and Big Sandy, MT. All three areas grow non-irrigated wheat. Non-irrigated winter wheat yields are used for Texas and Colorado while non-irrigated spring wheat yields are used for Montana with yields for both spring and fall plantings used for analysis. County level yields and prices from 1976 until 2008 (Figures B.1 and B.2) were collected from the National Agricultural Statistical Service (USDA-NASS, 2008).

The historical yields and prices were used to calculate gross annual returns for each location. Gross returns are adjusted for inflation to year 2005 using the Implicit Price Deflator for gross national product (USDC, 2009). Gross returns were also adjusted for direct government payments. It was assumed that the government base acreage was equal to the available production acreage on the farm. A five-year average yield was calculated as the base yield for the farm. The revenue from the direct payments was incorporated into the gross returns. Wolfley (2008) estimated farm-level operating costs for each of the three area. These costs estimates were used to calculate net-annual returns for each farm. To be consistent with the gross returns, the estimated costs were also adjusted to year 2005 dollars. The operating cost for Texas is $75.82, Colorado is $66.67, and Montana is $70.98 an acre.

The summary statistics for the three regions are found in Table B.1. The mean return for Montana is $127.73, Colorado is $105.35, and Texas is $70.33. As expected, Montana with the highest return also had the highest standard deviation (59.46). Surprisingly, Colorado (34.96) had the lowest standard deviation with Texas (44.37) falling in between the two states. The minimum and maximum returns for Texas and Montana reveal that both areas have the potential for high returns but also the potential for extremely low returns. Colorado on the other hand historically has not seen the large returns and losses. Locating production in Colorado from both the Texas and Montana perspective could provide some risk reduction based solely on the visual inspection of the summary statistics.

Included in the summary statistics is an analysis of the higher moments of the data, specifically the third and fourth moments. As expected from the summary statistics, Colorado had the lowest skewness and kurtosis. Texas had the highest skewness and kurtosis. Figure B.3 provides a visual method of evaluating the skewness and kurtosis of each production region compared to the normal distribution. The normal distribution has zero skewness and a kurtosis value of 3. All three regions have higher skewness and kurtosis than the normal distribution. Taking this analysis one step further, a normality test was done for each region (Table B.2). The normality tests confirm the fact that statistically, Colorado follows a normal distribution, but Texas and Montana do not. These results further motivate the need to go beyond linear correlation to measure the dependence between the three regions. The absence of normality justifies the additional effort of using copulas to model the joint distribution of returns from the three regions.

3.6 Results

The objective of this research is to evaluate the effectiveness of implementing asymmetric dependence structures and risk measures into a portfolio optimization problem. The discussion of results consists of two parts and addresses this objective. The first part is a discussion of the copula estimation. This consists of a discussion of which copula provides the best fit and a discussion of the different copula dependence measures. The second part of the results consists of a discussion of the actual portfolio optimization results including an analysis of the efficient frontiers using different dependence structures.

3.6.1 Copula Results

The first step in estimating copulas is to specify the marginal distribution for the returns from each region (Figure B.4). The Akaike information criterion (AIC) was used to determine the most appropriate marginal distribution. The Pearson5 distribution is the most appropriate distribution for the Texas returns. An Extreme Value Max distribution provided the best fit for both Montana and Colorado returns.

The fitted marginal distributions provide the necessary information to estimate the copulas. SIC, AIC, and HQIC criterion were used to select the most appropriate copula. The AIC has been shown to be the best for smaller number of observations. Based on the AIC, the Clayton copula provides the best fit, followed closely by the Frank and Gaussian copulas (Table B.3). Consistent across all three criteria is the fact the Gumbel copula does not provide a good fit which implies that there is little upper tail dependence. On the other hand, the Clayton copula fit measures indicate that there exists stronger lower tail dependence. Given these fit statistics, the rest of the analysis focuses on the Clayton, Frank, and Gaussian copulas.

Figure B.5 illustrates the difference when simulating from the Clayton, Frank, and Gaussian copulas. Each Figure represents 1,000 simulations from each copula. As expected, the Clayton and Frank have more dependence in the lower tails while the Gaussian has more dispersion throughout the simulated data. Figures B.6 and B.7 illustrate the simulated returns using the Clayton copula and multivariate normal relationships. The result of using the Clayton copula to capture the lower tail dependence is illustrated in the two Figures.

As discussed earlier, Kendall's Tau measure of rank correlation can be calculated using the copula estimation results. This provides a means of comparing 'apples to apples' when analyzing different dependency measures. Table B.4 consists of Pearson linear correlation and Kendall's Tau based on copulas and based on standard method. The Pearson correlation measure is greater than all the other methods. Focusing the comparison on the Rank Correlation measures, the copula based measures are greater than the standard method in all cases (Figure B.6). This result implies that the copula methods are capturing more of the dependence that exists between the individual returns. When comparing the results across copulas, the Frank and Gaussian Copula exhibit a stronger relationship between Texas returns and Colorado returns than does the Clayton Copula. The other relationships show similar magnitudes of rank correlation measures.

3.6.2 Efficient Frontiers

The efficient frontiers provide a starting point for examining the differences that occur if alternative dependence structures are incorporated into the portfolio optimization. The efficient frontiers for both asymmetric and normal dependence structures are illustrated in Figure B.7. The y-axis represents returns per acre and the x-axis represents the CVaR risk measure. The efficient frontiers illustrate the difference when asymmetric dependence is incorporated. Recall that the Clayton copula emphasizes lower tail dependence while the normal distribution treats the upper and lower tails equally. The normal frontier has higher levels of returns with lower levels of risk over all the points on the efficient frontier. Or in other words, by assuming non tail dependence, the multivariate normal approach underestimates the risk associated with a given level of return. This is consistent with the copula estimation results where the Clayton copula provides the best fit. Thus, by using the Clayton copula, more emphasis is placed on the downside risk. This is clearly seen in the efficient frontiers where for the same expected returns, the asymmetric frontier had a higher (more risky) CVaR level than the normal frontier.

3.6.3 Portfolio Allocations

The efficient frontiers provide a generalization of the possible risk efficient portfolios. The next step is to analyze the actual allocations for the three production regions. Figure B.8 illustrates the efficient allocations for the asymmetric and normal dependence structures. The vertical axis represents the percent of acreage allocation and the horizontal axis is the expected return. As expected, the level of allocation differs based on the expected return and this is consistent for both dependence structures. The acreage allocation shifts from all Texas to a combination of all three to finally all acreage in Montana as the expected return gets larger. When comparing the two different dependence structures, the main allocation difference occurs with Montana acreage becoming part of the efficient portfolio allocation sooner for the asymmetric dependence case. This is evidence of the fact that assuming a symmetric dependence structure is not capturing the tail risk that exists in Colorado. The asymmetric case takes that tail loss into consideration and for this reason incorporates acreage from Montana into the efficient acreage allocation.

Figure B.9 helps to separate out the allocations for the three regions. The acreage allocations for Texas are similar for both dependence structures. The Colorado allocations illustrate the difference that was also seen in Figure B.8. Under the normal dependence case, the acreage allocation for Colorado increases faster than the asymmetric dependence case. This is also seen in the Montana allocations where under the asymmetric case, acreage is shifted to Montana sooner than in the normal case. This illustrates once again the impact of including the lower tail dependence that is captured by the Clayton copula. By ignoring the lower tail dependence, efficient allocations would consist of more acreage in Colorado and would not use Montana to reduce the risk of the acreage allocation.

The results illustrate that ignoring asymmetric dependence structures is not a trivial matter. The copula fit statistics illustrate that the data used in this data does exhibit tail dependency and should not be ignored. The Clayton copula specification provides a method to capture the tail dependency that exists in the multivariate structure of the data. The efficient portfolio frontiers illustrate an applied case of not accounting for tail dependency. Using the CVaR as a measure of risk aversion, the acreage allocation and risk exposure is different for the asymmetric dependence structure versus the normal dependence structure. Across the entire efficient frontier, the normal approach had a lower level of risk for the same level of return when compared to the Clayton copula specification. In other words, by ignoring tail dependency, risk may be underestimated for a given level of expected returns.

3.7 Conclusions

As agriculture becomes more industrialized, the role of alternative risk measures and dependency structures will become more utilized. In this case, an asymmetric dependence structure along with a coherent risk measure (conditional value at risk) was applied to geographical diversification. In this case, geographical diversification relates to producing crops in three different regions. The land portfolio consisted of dry land wheat production acres in Texas, Colorado, and Montana.

Two of the three production regions were found to have non-normal returns distributions. Given this, two different dependency structures were used to model the data. A Clayton copula was used to capture lower tail dependence and was also found to provide the best fit based on Akaike information criteria. The second dependency structure was a Gaussian copula which assumes symmetric dependence in the upper and lower tails. These two measures were used to estimate the multivariate relationship between the three production regions. The portfolio optimization routine consisted of maximizing expected returns and minimizing the risk (CVaR). As expected, the risk exposure changes as the acreage is moved from the least risky region, Texas, to the most risky, Montana.

The implications of using an asymmetric dependence structure, or ignoring the asymmetric relationship, were also illustrated. The efficient frontiers under both assumptions showed that ignoring the asymmetric nature of the data could lead to optimal portfolio allocations that could underestimate the actual risk exposure. All optimal allocations under the symmetric dependence assumption had a lower risk exposure level than under the asymmetric dependence assumption. The implications of these results provide researchers with more motivation to move beyond the standard assumptions of linear correlation and normality.

The results of this report do not take into consideration the costs that could be involved with geographical diversification involving separate states. The significant distance between production regions means that issues such as transportation, equipment allocation and management of the production areas must be addressed. Each one of these topics could be topics for further research.

4. SCENARIO GENERATION FOR DISCRETE STOCHASTIC MULTI-STAGE PROGRAMS USING MOMENT MATCHING METHODS AND COPULAS TO MEASURE DEPENDENCIES

4.1 Introduction

The representation of uncertainty is a key element in a stochastic programming model (Topaloglou, Vladimirou, and Zenios, 2008a). This representation of uncertainty needs to be coherent and in a form that is suitable for computation. The formulation of a multi-period stochastic programming model exacerbates the problem of representing the uncertainty in a form that is computationally tractable because of the curse of dimensionality. The curse of dimensionality requires that multivariate continuous distributions or discrete distributions with a large number of outcomes be represented by a discrete approximation of the underlying distribution. Scenario generation is a method to obtain a discrete set of outcomes for the random variables. These set of discrete outcomes need to be small enough for computational ease but still represent the stochastic nature of the random variables. The random variables in this scenario generation framework are implemented into the optimization model as uncertain parameters. Random variables can include variables such as stock prices, commodity prices, yields, interest rates, and credit constraints, depending on the research problem. The discretization of both the random variables and the probability space result in a framework where the random variables take finitely many values. This also implies that the factors driving the risky outcomes are estimated via the generated scenarios. These generated scenarios can also be considered to be a sequence of events. Given this framework, the uncertainty in later time stages is captured by finitely possible outcomes for the next observation. This branching process of events is represented using a scenario tree. Each branch of the tree represents possible values of the random variables. An ideal scenario tree would represent the whole universe of possible outcomes of the random variables which would include optimistic and pessimistic projections.

Several different methods have been proposed to generate scenarios. Some of the common techniques used are time series econometric techniques such as vector autoregressive models and statistical methods including random sampling and bootstrapping. (Kaut, and Wallace, 2007). More recently, neural networks and copulas have begun to be used to generate scenarios (Kaut, and Wallace, 2009). The use of copulas to capture the multivariate relationship has been limited to single period problems and has not been extended to a multi-period application. The difficulty with time series, random sampling, and bootstrap techniques is that it has been shown that a large sample has to be drawn for the scenario generation technique to match the continuous distribution (Hoyland, and Wallace, 2001). In a multi-period setting, the curse of dimensionality limits the number of random samples and often hinders the effectiveness of these methods. An alternative to these methods that models the first four moments of the data and also the multivariate structure is a moment matching method (Hoyland, and Wallace, 2001). This method uses a non-linear optimization routine to minimize the statistical distance between the statistical moments of the generated discrete data and the statistical moments of the underlying distribution.

Previous uses of moment matching methods to generate scenario trees have relied on linear correlation to model the multivariate relationship between random variables. One of the drawbacks of linear correlation is that unless the random variables are normally distributed, the correlation that exists is not accurately measured. To overcome this weakness, copulas have been used in financial studies (Clemen, and Reilly, 1999; Dowd, 2005; Joe, 1997; Kallenberg, 2008) to capture the correlation that exists between random variables. The benefit of using copulas is that they model the marginal distributions of the random variable separately from the multivariate distribution. This allows the flexibility of modeling the marginal distributions in the most efficient manner without the same distribution for each random variable. The concept of using copulas to model scenarios has been applied to a single period setting but not to a multi-period application (Kaut, and Wallace, 2009).

The objective of this research is to extend the moment matching method of scenario generation by using copulas to capture the multivariate structure of the data. As mentioned earlier, copulas have been used to generate scenarios in a single period setting and through a random sampling method but have not been used in the moment matching specification nor in a multi-period setting. Incorporating copulas allows the flexibility of modeling the multivariate relationship beyond the normal assumption and the moment matching method allows the inclusion of skewness and kurtosis. A farm geographical diversification problem is used to illustrate the application of the moment matching scenario generation methodology. The optimization problem consists of a discrete stochastic programming (DSP) model that is formulated to optimize a farmer's portfolio, where the portfolio is the acreage allocations between three distinct geographic areas.

This section is organized as follows. First, previous literature concerning scenario generation, multi-stage optimization, and copulas are presented. Second, the methodology of the scenario generation and copulas are outlined. Third, is a discussion concerning the DSP model, including a discussion of the objective function and constraints. Fourth is a discussion of the moment matching results, copula fitting results, geographical diversification results, and conclusions.

4.2 Literature Review

Agriculture is a natural application of sequential decision problems. The current decisions that producers make have implications on future actions. These decisions are made with uncertain knowledge concerning the future, complicating the decision process (Kaiser, and Boehlje, 1980). As stated by Hardaker et al. (1991), "Uncertainty is important because it affects the consequences of decisions in ways that decision makers are not indifferent about." Currently, agricultural producers are facing a myriad of future uncertainties. These include but are not limited to the following: changing governmental policies, highly volatile commodity prices, inflation rates, interest rates, and rising production costs. Exacerbating this problem further is that these uncertainties need to be considered for multiple years during the planning process (Featherstone, Preckel, and Baker, 1990; Hardaker, Pandey, and Patten, 1991; Torkamani, 2005). Large capital purchases that generate returns over multiple periods and multi-year land rental contracts and production contracts require that agricultural producers develop investment and production strategies beyond one year.

Unlike in a traditional finance setting, changing an agricultural portfolio (crop mix) does not consist of the simple mechanism of buying and selling certain stocks (Robison, and Barry, 1980). Capital assets in agriculture are often illiquid and are fixed in the short-term. This situation, referred to as asset fixity, complicates the reallocation of crops resources (Edwards, 1959; Johnson, and Pasour, 1981). Precise crop rotations, high capital costs, and costly entry and exit conditions into certain crops require that producers formulate planning strategies for multiple periods.

Agricultural economists have tackled the problem of dealing with both the uncertainty and multiyear planning periods by using discrete stochastic programming (DSP). Rae published two seminal articles concerning the application of discrete stochastic programming in agriculture (Rae, 1971a; Rae, 1971b). In his first article, he examined a three-period fresh vegetable operation. The states (random outcomes) were defined both by predefined weather conditions and crop prices. He noted that one of the inherent weaknesses with DSP was the lack of more states or values for the stochastic variables. The argument against more states was based on the “curse of dimensionality”. There is a tradeoff of complexity and solvability when using dynamic programs. Rae’s second paper dwelt with viewing the problem using Bayesian decision theory. He also investigated the use of alternative utility functions within the objective function. He concluded that the ability of discrete stochastic programming to handle alternative utility functions makes itself a useful tool when studying sequential decision problems in agriculture.

Discrete stochastic programming has been used to model multi-stage wheat marketing (Kaiser, and Appland, 1989; Lambert, and McCarl, 1989), fixed versus adjustable rate loans (Leatham, and Baker, 1988), capital structure (Featherstone, Preckel, and Baker, 1990), assess prospective technologies and organic farming (Flaten, and Lien, 2007; Torkamani, 2005), and economic efficiency (Torkamani, and Hardaker, 1996). In finance research, discrete stochastic programming is now used to model asset allocation and for portfolio optimization routines (Hochreiter, 2010; Infanger, 2006; Lien, et al., 2009; Mulvey, and Shetty, 2004; Topaloglou, Vladimirou, and Zenios, 2008b).

The topic of scenario generation for multi-stage stochastic programming has received very little attention in the agricultural economics literature. Featherstone et.al (1990) discussed different methods of generating scenarios. The method they used to generate the scenarios consisted of partitioning the joint probability distributions into a selected number of regions. The conditional mean and probability for each region of the partitioned joint probability distribution was estimated. This method maintains the mean of the distribution but has been shown to underestimate the variance (Miller and Rice, 1983). Leatham and Baker (1988) used Monte-Carlo simulation to generate random stochastic values. They then ranked the simulated values and defined a high and low value. This method ignores the statistical characteristics of the underlying data (moments) and would be likely to misrepresent the actual stochastic nature of the random variables.

Previous research has noted that there are three key elements to discrete stochastic programming: (1) specification of the objective function, (2) definition of the constraints, and (3) definition of the random variables (Featherstone, Preckel, and Baker, 1990). This research relies on the previous definitions of the objective function and constraints provided by Featherstone, Preckel, and Baker, and focuses on the third element, definition of random variables.

Inherent within the third element, defining the random variables, are three issues that must be addressed when representing a stochastic process with a finite number of generated scenarios. Some authors have often referred to this representation as a deterministic equivalent of the underlying stochastic process.(Cornuejols, and Tutuncu, 2009). The first issue is modeling the correlation over time. An autoregressive modeling structure is used to capture the autocorrelation between the variables. The second issue is modeling the univariate and multivariate structure of the random variables. The univariate structure is modeled using the moment matching technique which captures the first four moments. The multivariate structure is modeled using copulas. The third issue is determining the appropriate size of the scenario tree. The number of scenarios needs to be large enough to accurately represent the stochastic process but small enough to allow the program to be solvable. The next section discusses each of these three issues in more detail.

4.3 Scenario Generation

In a multi-stage stochastic program, the decisions can be made at multiple points in time, referred to as stages. In a production agriculture framework, the stages would represent the annual cropping decision. Let $n \geq 2$ represent the number of stages. The random events that occur at each stage is represented by ω which is a vector$(o_1, \cdots, o_{n-1})$. The first stage decisions are made before any component of ω is revealed. Once the first stage decisions are madeo_1 is revealed and then the second stage decisions are made. After that, o_2, is revealed and the same pattern continues on. A scenario tree is often used to better illustrate the multi-stage decision framework.

4.3.1 Scenario Trees

A scenario tree is shown in Figure C.1. A scenario tree is defined by its nodes, and branches. The nodes represent the states of nature at a specific point in time and are labeled 1 through 15 in Figure C.1. Each node also corresponds with a specific stage or time stage. In Figure C.1 these stages are labeled as *t with t = 1,…, T*. Within the scenario tree there are three different types of nodes. The root node, node 1, represents the initial stage or 'today' and is immediately observable from deterministic data. There is only one root node per scenario tree. Each node is in one stage and each node i in stage $k \geq 2$ is adjacent to a unique node $a(i)$ in stage k-1. The node $a(i)$ is referred to as the *father* of node i. For example, the father of node 5 is node 2, or $a(5) = 2$. Leaf nodes are the final nodes in the scenario tree. These nodes do not have any successors that follow them. The actual scenario consists of the paths from the root node to the leaf nodes. Thus the actual number of generated scenarios is equal to the number of nodes in the last stage (8 in Figure C.1). In between the root and leaf nodes are the intermediate nodes. In this case, decisions will be made at the root and intermediate nodes. Each branch of the tree represents a possible value of the random variable which implies that there is a probability of occurrence for each node. At each stage, the sum of the probabilities for the nodes sums to 1. An ideal scenario tree would represent the whole universe of possible outcomes of the random variables which would include optimistic and pessimistic projections.

4.3.2 Scenario Tree Generation

The scenario tree provides a visual representation of the generated stochastic data. The scenario generation approach using sequential optimization requires that the statistical properties of the random variables be specified. The scenario tree is then constructed so that these pre-specified statistical properties are satisfied. These properties are maintained by letting the stochastic variables and probabilities are decision variables in a non-linear optimization problem. The objective function in the non-linear problem is to minimize the square distance between the specified statistical moments and the new generated statistical moment. This minimization occurs at each node in the scenario tree. This implies that the branches (and the associated random values for each branch) emanating from each node represent the statistical moments of the underlying distributions of the constructed scenario tree (Hoyland and Wallace, 2001). The optimization method applied in this research is referred to as a sequential optimization (Gulpinar, Rustem, and Settergren, 2004). The sequential optimization implies that at each stage and node, the distance between the historical statistical properties and scenario generated statistical properties is minimized. The historical data is then updated with the scenario generated data and the historical moments are re estimated accounting for the new data. The optimization method is then repeated for each node and stage.

Following the notation of Hoyland and Wallace (2001), the set of all specified statistical values is denoted by S and the value of the specified statistical value i is SV_i, for all i that exist in S. In this application, the statistical values in the set S consist of the first four moment of the historical distributions and the dependence measure between each random variable (thus $i = 1,...,5$). The random variables and probabilities are denoted by x and p. In this application, x would represent the random land prices and dry land wheat returns. The mathematical expression of the statistical property i can be defined as a function of the random variable x and probability p, $f_i(x,p)$. In this application $f_i(x,p)$ is used to define the first four central moments and the dependence measure between i and j[2]. The objective is to construct x and p so that the sum of square deviations between the historical statistical properties and the statistical properties of the constructed distributions is minimized. In its most general form, a scenario generation model can be specified as:

$$\min_{x,p} \sum_{i \in S} w_i (f_i(x,p) - SV_i)^2 \quad (4.1)$$

$$s.t. \quad \sum p_i = 1, \; p \geq 0, \quad (4.2)$$

where w_i is the weight of statistical property i and the other variables are the same as previously defined. Equation (4.2) is the probability constraint which forces the

[2] The specifics of the dependence measures are discussed in Subsection 4.2.3.

probabilities to be greater than zero and sum to one. A more detailed model is provided below.

The non-linear problem specified above is often not convex, which implies that the solution may be a local optimum and not a global optimum. In many cases this is not satisfactory, but in the case of scenario generation, the local optimum is sufficient (Gulpinar, Rustem, and Settergren, 2004). An objective value close to zero indicates that the constructed distribution has a good match to the underlying distribution. The weights (w_i) can be used to incorporate relative importance to the statistical specifications.

The advantage of using an optimization approach is that any central moments and co-movements can be part of the statistical specifications of the distribution and implemented into the objective function found in equation 1(Hoyland, and Wallace, 2001). The first four moments will be considered in this study. The dependency between variables will be modeled using copulas which are discussed in the next section. Let I = (1,2,...n) denote the set of random variables. Let M_{ik}, , for k = 1, 2, 3, 4, be the first four central moments of the historical distribution of random variable i. The dependence between random variable i and l (such that $i, l \in I$ and $i < l$) is denoted by D_{il}. Let B_t be the number of branches from a node at stage $t = 0, \ldots, T-1$. The scenarios x_{ij} for random variable $i \in I$ and probabilities p_j for $j = 1, \ldots, B_t$ of the historical distributions are decision variables in the following non-linear optimization problem:

$$\min_{x,p} \sum_{i=1}^{n} \sum_{k=1}^{4} w_{ik}(m_{ik} - M_{ik})^2 + \sum_{i,l \in I, i<l} w_{il}(d_{il} - D_{il})^2 \tag{4.3}$$

$$s.t. \quad \sum_{j=1}^{B_t} p_j = 1, \tag{4.4}$$

$$m_{i1} = \sum_{j=1}^{B_t} x_{ij} p_j, \quad i \in I, \tag{4.5}$$

$$m_{ik} = \sum_{j=1}^{B_t} \left(x_{ij} - m_{i1}\right)^k p_j, \quad i \in I, \;\; k = 2,3,4, \tag{4.6}$$

$$d_{il} = C\big((F_i(x_i), F_l(x_l)\big), \qquad i, l \in I \text{ and } i < l \tag{4.7}$$

$$p_j \geq 0, \;\; j = 1, \dots, B_t, \tag{4.8}$$

where w_{ik} and w_{il} are weights which capture the relative importance of the central moments and dependence of the random variables $i, l \in I$. The first half of the objective function (4.3) represents the minimization of the square norm's distance between the scenario constructed moments (m_{ik}) and the historical moments (M_{ik}). The second half of the objective function represents the minimization of the norm's distance between the historical dependence measure (D_{il}) and the scenario constructed dependence measure (d_{il}). In this case, the dependence is measured by the Gaussian copula. Thus the first half of the objective function matches the moments and the second half maintains the multivariate structure of the random variables. The first constraint shows that the probabilities must sum to one at each branch. The rest of the constraints are used to formulate the first four central moments of the variables and the copula dependence measure. The last constraint is to ensure that probabilities are non-negative.

It is important to note that the estimated moments of the distributions are conditional on past history and are conditional on the associated path of the scenario tree. This implies that the historical data is updated with the new scenario generated observation after each scenario estimation. Thus the updated historical moments of each distribution are estimated using both the historical observations and the new observations generated through the scenarios.

4.3.3 Copulas

As noted above, copulas are used to model the dependence that exists between the random variables. The standard method to measure the dependence measure is the Pearson's correlation which is often referred to as the linear correlation. Copulas are often the preferred method to capture this dependence because they are able to capture more than just the linear dependence because they allow the flexibility of separating the marginal distributions from the joint distribution. An understanding of the origin of the word copula helps to explain better their usefulness in modeling joint distributions. The word copula is derived from the word couple, or in other words, copulas couple the marginal distribution to the joint distribution. The benefit of the flexibility that copulas allow is that marginal distributions can be modeled separately from the joint distribution. For example, under the assumption of multivariate normality, it is assumed that all marginal distributions are normally distributed. If that is true, then there exists no problems, but as is often the case, the assumption of normality for the random variables may not be appropriate(Just, and Weninger, 1999). Copulas allow the marginal distributions to be modeled individually and the copula function will join the individual marginal distributions to the joint distribution. The evidence of the advantages of using copulas to measure dependence can be found in finance and statistics journals, and recently in the agricultural economics literature (Bai, and Sun, 2007; Clemen, and

Reilly, 1999; Joe, 1997; Patton, 2002; Rank, 2000; Trivedi, and Zimmer, 2005; Vedenov, 2008; Xu, 2005; Zhu, Ghosh, and Goodwin, 2008). An extensive treatment of copulas can be found in numerous books and research articles (Patton, 2002).

The origin of copulas can be traced back to the Sklar theorem (Sklar, 1959). The Sklar theorem allows one to construct joint distribution of several random variables based on their marginal distributions and a copula. By definition there are an infinite number of copula functions, therefore an infinite number of joint distributions that may be generated for given marginal distributions. Various copula families have been used in risk research (e.g. Gaussian, Archimedean, etc. (Hennessy, and Lapan, 2002)). The Gaussian copula fit the data the best (discussed in results section) and is used as the dependence measure[3].

The copula based joint cdf is obtained by transforming the marginal distributions to standard uniform distributions. One can view this joint cdf as the joint distribution stripped of all information about the marginal distributions. The only thing remaining is the information about the joint distribution multivariate structure. Therefore, copulas enable the decoupling of the marginal distributions from the multivariate structure. This gives the modeler much more flexibility in modeling multivariate relationships. In this study, the marginal distributions are modeled using an empirical distribution. This does not enforce any assumed distributional form on the marginal distributions.

---

[3] Five different copulas were analyzed, Gaussian, T, Gumbel, Frank, and Clayton. The Gumbel, Frank, and Clayton are all part of the Archimedean copula family (see Nelsen (2006) for detail concerning each copula). An illustration of the shapes of the different copulas is shown in Figure C.6. One of the advantages of using copulas is that more emphasis can be placed on the tails of the data and this is clearly evident with the Clayton and Gumbel copulas. The T-copula has equal tail weights and so no emphasis can be placed on the lower tail of the data. One of the main disadvantages of the Gaussian copula is that it places no weight on the tails of the data.

Implementing copulas into the scenario generation process consists of two parts. The first part consists of estimating the bivariate copula for each pair of data. The parameters for the copula are estimated from the historical data.

The next step consists of generating the actual scenarios. The scenario generation method described above is used to generate the scenarios implemented into the mathematical model. At each time period, a specified number of nodes are used to estimate the branches on the scenario tree and its associated probability.

4.4 Farm Problem Specification and Data

Traditional farm diversification problems have focused on crop mix in a given production region. The optimal crop mix consisted of one that maximized returns while minimizing the risk in the traditional portfolio optimization framework. This method of viewing diversification via crop mix has been the standard since the 1950's (Heady, 1952). Changes in agriculture markets and the investments in agricultural land from large investment firms motivate a need to look beyond the traditional crop mix diversification approach. Geographical diversification is a method being used by some producers and land investors to manage risk. Often, these producers are highly specialized and focus production on one specific commodity. The opportunity to diversify risk if production takes place in only one region is limited. Geographical diversification provides an opportunity to diversify away some of the production risk associated with a given commodity by taking into consideration weather patterns, disease, and pest problems that are not homogenous among the regions.

4.4.1 Capital Allocation Model

This type of asset allocation model can be viewed a multi-stage dynamic decision problem. The decisions take place at discrete time points. At each decision point the farmer has to evaluate the previous stage's market conditions and the composition of the enterprise diversification. At the same time, the farmer must evaluate future conditions such as future yields and prices. All this information is then used by the farmer to reallocate or adjust how the land is allocated in three regions. This could involve increased short term or long term borrowing because of increased operating expenses, machinery purchases, and land purchases. This same decision process continues through the decision stages of the model.

4.4.2 Specification of Objective Function

At the beginning of each decision stage, the farm manager is faced with many important decisions. Once a farmer makes a decision on land or crop allocation, it is often very costly and difficult to rearrange the land allocation. Some of those decisions are the levels of investment in farmland, capital purchases such as machinery to service new crops or acreage, and debt financing on farmland and capital. These decisions are not limited to one decision stage but will be made over a finite horizon planning stage. The decision process is even more difficult because these allocation decisions are based on the realization of uncertain events. Because of this uncertainty, the farm manager's objective in making these decisions is to maximize expected utility subject to land and capital constraints. Specifically, for this problem, the farm manager is seeking to maximize the expected utility of terminal net wealth. This model specification follows the specification developed by Featherstone, Preckel, and Baker (1990) for a stochastic dynamic programming model. The objective function is defined as

$$\max_{\{x_{i,j,t}\}_{t=0}^{T}} E_0[u(W_t)], \tag{4.9}$$

for all T.

Equation (4.9) specifies the objective function of this model. The objective of this model is to maximize the expected utility of ending owner equity, where ending owner equity is used as the measure of terminal net wealth. Ending owner equity is calculated as the income from production activities, less depreciation, interest expense, and family living withdrawals, plus the previous year's owner equity at the terminal time period T. Maximization of terminal net wealth is used because of the difficulty of implementing an additive utility function (Featherstone, Preckel, and Baker, 1990). An additive utility function assumes that there is independence between periods. In reality, the assumption of independence is often not the case, so terminal wealth is used to avoid that problem. A discussion on the appropriateness of functional form for the utility function has been provided by both Rae (1971) and Featherstone , Preckel, and Baker (1990). The non-separable negative exponential utility function (Torkamani, 2005) is used and defined as

$$U(w) = 1 - exp[-\{(1-\lambda)r_{min} + r_{max}\}w], \qquad for\ 0 \le \lambda \le 1, \tag{4.10}$$

where λ is the risk preference parameter with r_{min} and r_{max} are the upper and lower bounds of the coefficient of absolute risk aversion (r_a) with *w* being the ending wealth. Torkamani (2005) provides the relationship between the coefficient of risk aversion, r_{min}, r_{max}, and λ by specifying the following function:

$$r_a = (1-\lambda)r_{min} + r_{max}, \quad for\ 0 \leq \lambda \leq 1.$$

The range of r_a is also taken from the results derived by Torkamani (2005).

4.4.3 Constraints

The constraints in this model consist of three types, land, machinery, and financial constraints (See appendix A for a mathematical formulation). The first land constraint limits the amount of land sold or farmed to be equal to or less than the amount of land owned. The second land constraint transfers the land from one stage to another. The third land constraint limits the total amount of land both rented and owned in each region to a specified amount. This amount is given an upper bound and lower bound. The upper bound of land used in this research is 5,000 acres. It is assumed that the farmer can feasibly operate this amount of acreage.

The purpose of the machinery constraints are to ensure that there is enough machinery capacity to serve the acreage operated by the farmer. The first machinery constraint sets the amount of machinery owned to be greater than the necessary amount to farm the owned and rented land. The second machinery constraint transfers machinery from one stage to the next. Though this constraint sounds very simple, just as with the land transfer, it is necessary the owned machinery value and capacity is properly transferred to subsequent stages. The final machinery constraint sets the amount of machinery that needs to be purchased based on increases in acreage both from increased owned and rented land.

There are three type of financial constraints used in the model. Because funds available from borrowing are not limitless, a constraint is used to limit the amount of borrowing. The next constraint is a balance sheet constraint that enforces the standard accounting equation of total assets equal total liabilities plus owner's equity. The final constraint calculates the change in owner's equity and transfers it to the subsequent stages. The transfer of equity occurs until the final or terminal node which represents the end of the planning horizon. At this point, the ending owner's equity is the value used to measure wealth in the objective function.

4.4.4 Farming Situation

Prior to solving the dsp model, the base units needed to be specified for the representative farm and financial coefficients used in the model. The initial owner's equity is $1,100,000, with initial debt of $500,000, and assets of $1,600,000. The transaction cost of selling land is assumed to be the standard 6% and for the machinery a 15% level is used. The 15% represents the additional cost that is incurred if machinery is sold or bought. The dsp model was solved using various starting points and scenarios. One assumption that maintains fixed across all models is that the risk aversion parameter remains the same. It is assumed that the farmer is risk averse and the RAC (risk aversion coefficient) is set at 0.05 (McCarl, and Bessler, 1989). The model was first solved assuming that geographical diversification was not an option. The production was constrained to the base region. The model was solved three times, varying the base region between Texas, Colorado, and Montana. The model was then solved assuming that the farmer could geographically diversify. Once again, there was a base region and the farmer could allocate production to either the base region or the other two regions. The model was once again solved three times varying the base region between Texas, Colorado, and Montana. The base region began with the most acres and could not drop below a specified level of 3,000 acres. The results of each are discussed below.

4.4.5 Land Data

This research is concerned with growing dry land wheat in three potential growing regions. The three different growing regions represent opportunities to diversify certain elements of production risk inherent in agriculture. Similar to the classic stock portfolio case, the goal is to find the optimal allocation of land in the three regions. The data necessary to analyze this problem are the actual net returns (gross revenue less cost of production) and the land prices for each region.

Land is often not included as an investment instrument in the traditional farm portfolio analysis, but land is a large part of a farmer's balance sheet and plays a pivotal role in obtaining credit. As was seen in the 1980's (Figure C.2) land values dropped dramatically causing a crisis to occur in the agricultural lending industry. More recently, the concern has been over the rapid increase of land prices. In a period of three years, 2005-2008, land prices in Texas increased by 87%, Colorado by 39%, and Montana by 88%. High commodity prices, decreasing supply of farmland due to urbanization, and increased interest in farmland investments have driving much of the high demand of farmland. The summary statistics of land prices are found in Table C.1. Colorado had the highest mean value of land, but the value of land has not increased as fast as Montana and Texas. Montana has seen the greatest change in farmland value. This particular county in Montana is particularly known for consistently high wheat yields and the high wheat prices have been a large driver of the increase in land value. The mean increase in farmland value for the three regions over the past thirty years is $16 for Texas and Colorado, and $24 for Montana (Table C.2). Land prices in all three regions exhibit skewness of greater than one which implies a departure from normality. The kurtosis for a normal distribution is approximately three. All three areas are greater than three and exhibit the need to account for kurtosis. A histogram and pdf's of land prices for the three regions illustrates the shape of the data (Figures C.3 and C.4). All three regions exhibit the skewness and kurtosis inherent in the data with more weight being placed on the lower tails of the data. The shape of the land prices illustrates the importance of not imposing the distributional assumption of normality. The skewness and kurtosis exhibited in the data would be underestimated in the lower tail and overestimated in the upper tail.

To account for the time series component that exists within land prices, an autoregressive model is used and is based on the specification provided by Lohano and King (2009):

$$P_t = \alpha_0 + \alpha_1 P_{t-1} + \alpha_2 R_{t-1} + \varepsilon_{2t}. \qquad (4.11)$$

After the first pass of estimations, it was found that the lagged values of gross returns (R_{t-1}) were not statistically significant, so the equation was re-estimated using OLS with the dependent variable as current land prices and the independent variable was lagged land prices and not lagged returns (Table C.3). Each lagged term for Montana, Colorado, and Texas were found to be significant and the models illustrate that current land prices are explained by previous period land prices. The errors terms also satisfy the condition that they follow a white noise process (Table C.4).

4.4.6 Gross Returns Data

Gross returns are calculated for each region using county level yields and prices gathered from National Agricultural Statistics Service and covered the time period of 1974-2008(USDA-NASS, 2008). Gross returns were also adjusted for direct government payments. It was assumed that the government base acreage was equal to the available production acreage on the farm. A five-year average yield was calculated as the base yield for the farm. The revenue from the direct payments was incorporated into the gross returns. Montana had the highest gross returns and also the highest standard deviation (Table C.1). Colorado had the lowest standard deviation among the three production regions. One note of caution with analyzing gross returns was the high returns that occurred in 2007 and continued in 2008 because of the record prices for commodities. For example, the net returns, from 2005 to 2008, increased by 120% in Texas, 107% in Colorado, and 73% in Montana (Figure C.5). All three regions also exhibit a non-normal shape (Figures C.3 and C.4) which is accounted for by including the third and fourth moments into the scenario generation routine.

Unlike land prices, gross returns had no clear time component. Although there were large deviations from 2005 to 2008, the previous time periods did not exhibit a large change. For this reason, the stochastic nature of the data is modeled by using the underlying distributions. The statistical characteristics of the underlying distributions for returns from the three areas were estimated and used in the moment matching estimation.

The final step was to analyze the normality of the data. This was done using the Shapiro-Wilk normality test. The results of the test are shown in Table C.6. The null hypothesis that the returns and land residuals were normally distributed was rejected for Montana land residuals, Texas returns, and Colorado returns. The results of this test are further validated by looking at skewness/kurtosis test (Table C.7). The test shows that the Montana land residuals, Texas returns, and Colorado returns all exhibit statistically significant skewness and kurtosis. Ignoring the higher moments of the data would not capture the 'whole story', thus ignoring possible risks.

4.4.7 Correlation Analysis

The correlations between the different production regions help to establish a preliminary hypothesis concerning the optimal allocation (Table C.8). Following the traditional portfolio theory approach, negative correlation is one indicator that implies risk reduction through diversification while positive correlation implies less risk reduction benefits from diversification. Net returns from Montana are negatively correlated with Texas and Colorado land prices. They are also negatively correlated with Texas returns. This would suggest that a Montana producer would benefit from diversifying production to Texas and vice versa. In addition, using variance as a risk measure, Montana had the highest variance of returns, while Colorado had the lowest. This implies that Montana could reduce its risk (variance) by allocating production acreage to these other regions. The same scenario would hold true for Texas as well by diversifying land to Colorado. Colorado land on the other hand is positively correlated with everything except Montana returns. This would imply that Colorado may not find it beneficial to incorporate geographical diversification to increase wealth and decrease risk. The results of the optimization model provide details concerning the optimal acreage allocations and are discussed in the next section

4.5 Empirical Results

The results of any math programming model are only as good as the input data. The discussion of the results begins with a discussion of the scenario generation results. This includes a discussion of the copula fitting and actual generated scenarios. This is followed by a discussion of the capital allocation model results.

4.5.1 Copula Results

Five different copulas were fit to the data to determine which one would best model the multivariate relationship. The five copulas were Gaussian, T, Frank, Gumbel, and Clayton copulas. As discussed earlier, the Clayton and Gumbel copula place more emphasis on the lower (Clayton) and upper (Gumbel) tails of the distribution. Figure C.6 provides an illustration of the shapes of the different copulas. The other three assume symmetric dependence between the upper and lower tails. Using an AIC fit criteria, the Gaussian copula provided the best fit for eleven of the sixteen bivariate relationships (Table C.9). Based on these results, the Gaussian copula is used to model the bivariate relationships and is used in the moment-matching scenario generation routine as the measure of dependence.

4.5.2 Scenario Generation Results

The goal of scenario generation is to generate a discrete set of data that captures statistical characteristics of the underlying data. In this case, in order to avoid any distributional assumptions, the discrete set of data is modeled to match the first four moments of the underlying data and also use the Gaussian copula to model the dependence. Five nodes are used as the discrete outcomes for each stage with five total stages. Five nodes were chosen to maintain the computational tractability of the problem. This implies that there are five nodes in stage one, 25 nodes in stage 2, 125 nodes in stage 3, 625 nodes in stage 4, and 3,125 nodes in stage five. Thus the model is solving over a total of 3,905 nodes. An illustration of the first two stages can be found in Figure C.7. The generated data for each node and the associated probability of the node are shown in the Figure. Figure C.7 helps to visualize what he moment matching routine is accomplishing over the planning stage. The statistical properties of the generated data compared to the underlying data are shown in Table C.10. As shown in the table, the moment matching method closely fit the first four moments and thus capturing the statistical nature of the underlying distributions.

4.5.3 DSP Results

The results for the optimal acreage allocations when production is limited to one region are shown in Table C.11. The assumption is that the farmer begins with 3,000 acres and has the opportunity to expand acreage up to 10,000 acres all in the same region over the defined planning period. The diversification strategies and optimal wealth levels (Figure C.8) of Texas and Colorado are similar. Both regions increase acreage in the beginning stages and hold the acreage roughly constant for the final two stages. Texas acreage is slightly greater than the allocated Colorado acreage and both Texas and Colorado acreage is greater than Montana acreage. Montana acreage allocation consists of increasing acreage in the first stage and then slowly decreasing acreage over the remaining four stages. Montana had the highest expected wealth level of the three regions. All three areas also show the standard deviation between the optimal allocations increase over the planning stage as well. This can be explained by the fact that the model is trying to optimize over 3,125 terminal nodes and over multiple stages.

The optimal allocation results incorporating geographic diversification are shown in Table C.12. When Texas is the base region, acreage is purchased in all three regions in the first stage. The acreage in Montana slowly decreases over the remaining stages while the acreage in Colorado and Texas remains fairly constant (see Figure C.9). There is a positive increase in the certainty equivalent of wealth for Texas when diversification is implemented (Figure C.8). When Colorado is considered to be the base region, different optimal diversification strategies arise. The land in Colorado will vary only slightly over the planning period with a slight decrease in the beginning but increasing to the initial acreage level in the final period (Figure C.10). The largest amount of land will be purchased in Texas with roughly two thousand acres of production over the entire planning period. Land in Montana on the other hand will increase in the beginning but be reduced to the initial acreage level in the ending stages (Figure C.10). The results when Texas is the base region and when Colorado is the base region provide a consistent diversification strategy and both scenarios provide an increase in the certainty equivalent of wealth. Both regions show that there is a beneficial diversification strategy that involves Texas and Colorado but Montana does not provide the same benefit.

The last scenario to consider is when Montana is the base region. Surprisingly, the optimal diversification strategy consists of the majority of allocation to be in Texas (Figure C.11). Part of this can be explained by the fact that Montana land prices are negatively correlated with Texas land prices and Texas returns. Montana returns and Texas returns are positively correlated but at a level which suggests that the returns from the two regions are close to being statistically independent. This is unlike the two previous scenarios which maintained the largest amount of acreage in the base regions. In the initial stages, additional land will be allocated in all three regions, but then acreage will be reduced in Colorado and Montana but not Texas. Thus a Montana farmer can increase certainty equivalent of wealth by transferring acreage from Montana to alternative production regions. This can partially be explained by the high volatility of dry land wheat returns in Montana. Both Colorado and Texas had lower returns, but also lower volatility. Part of this can also be explained by the assumption of a risk averse producer. A risk averse producer would be more concerned with volatility than the higher expected returns.

These results illustrate an important fact concerning geographic diversification. There is no one optimal allocation when dealing with geographic diversification. The optimal allocation is dependent upon the assumptions concerning base acreage and feasible locations. Another consistent theme across all scenarios is that the benefits of geographic diversification are positive and that as commercial agriculture expands, geographic diversification will become a more relevant diversification strategy.

4.6 Conclusions

A multi-period discrete stochastic programming model was formulated to analyze geographical diversification. Specifically, it analyzed whether a farmer would expand by buying more land locally or expand to other regions. The production of dry land wheat consisted of three different regions: Texas, Colorado, and Montana. The objective function consisted of maximizing terminal net wealth. The model analyzed the decision of how a farmer would allocate land to different production regions. Land is one of the most important resources a farmer has. Land traditionally composes a large share of the farmer's balance sheet. It is the base for loan collateral and future wealth. Not only is it important to consider the revenue stream from production on the land but also returns from land appreciation. The inclusion of both aspects is critical to effectively model geographical diversification decisions.

Discrete stochastic programming models both land prices and production revenue in a dynamic setting. As a farmer looks to make large investments in land and machinery, it is important to consider the results of the investment over multiple periods and not just look at the single period consequences. Discrete stochastic programming breaks away from the single period methodology of the traditional portfolio optimization and analyzes the optimal investments in a dynamic setting.

This research introduced a new method to generate the scenarios used in the dsp model. A moment matching routine was developed and the multivariate relationship between the random variables was captured using copulas. The first four moments of the underlying distributions were modeled. A Gaussian copula provided the best fit and was used as the appropriate dependence measure. The use of copulas provides a better method to estimate the dependence between the random variables when they are not normally distributed. Copulas are able to capture the non-linear dependence that may exist between two random variables and thus more accurately represent the relationship.

The results of this research also indicate that there are possible gains from geographical diversification. Wealth levels are increased for all three regions when production is diversified over the different regions. The optimal allocation of land to alternative production regions was dependent upon the base acreage assumption. One important factor of geographic diversification that needs to be considered is the additional costs incurred. Future research could take into consideration not only the wealth benefits but also the additional management, transportation, and labor costs that may occur.

5. SUMMARY AND CONCLUSIONS

Diversifying yield risk through geographical diversification has not been heavily addressed. Expanding beyond traditional enterprise diversification enhances the decision maker’s available risk management tools. This research has taken a closer look at the topic of geographical diversification. This research will provide a foundation for understanding the dependence of wheat, cotton, and sorghum yields on a broad geographic scope. In addition, not relying solely on linear correlation will enhance the understanding of yield relations. The implementation of copulas allowed the estimation of alternative dependence measures. Copulas also provided a method to estimate the tail dependence coefficient. Often times it is the tails of the data that are of concern. Copulas allow those regions of the distribution to be analyzed.

This research also looked at the implementation of coherent risk measures. Efficient risk management relies on good risk measures. The conditional value at risk (CVaR) measures overcomes many of the limitation of the traditional value at risk (VaR) measure. Using CVaR as the risk measure, geographical diversification was examined using portfolio optimization. The results indicated that the diversification of the portfolio was sensitive to not only the risk measure but also to the dependence measure. These results provide a starting point to begin analyzing agricultural problems using these new techniques.

Geographical diversification was also examined in a multi-period framework. There are some additional complexities when transitioning from a single period model to a multi-period model. One of the biggest hurdles is representing the historical data in a form that keeps the model manageable. Scenario trees have become a tool that provides a method to accomplish this. The stochastic path of the data is represented in a discrete form in a scenario tree. A new method to generate the scenarios was formulated. The method consisted of a non-linear optimization that matched the first four moments of the distribution and also implemented the use of copulas to model the multivariate relationship. The generated scenarios were then used as inputs to solve a discrete stochastic programming model. The results show that geographical diversification may have some benefit to farmers. The benefit is dependent upon where the initial acreage is allocated.

This research builds on the strong foundation of risk diversification that has been laid by agricultural economists. The goal is to add to the existing structure by incorporating these new techniques to old problems. These results will provide a backdrop to future research into the portfolio optimization problems of agricultural enterprises.

REFERENCES

Accioly, R., and F. Y. Chiyoshi. 2004. "Modeling Dependence with Copulas: A Useful Tool for Field Development Decision Process." *Journal of Petroleum Science and Engineering* 44(1): 83-91.

Acerbi, C. 2002. "Spectral Measures of Risk: A Coherent Representation of Subjective Risk Aversion." *Journal of Banking & Finance* 26(7): 1505-1518.

---. 2007. "Coherent Measures of Risk in Everyday Market Practice." *Quantitative Finance* 7(4): 359-364.

Alexander, G. J., and A. M. Baptista. 2002. "Economic Implications of Using a Mean-Var Model for Portfolio Selection: A Comparison with Mean-Variance Analysis." *Journal of Economic Dynamics and Control* 26(7-8): 1159-1193.

---. 2003. "Cvar as a Measure of Risk: Implications for Portfolio Selection." Paper presented at Paper Presented at EFA 2003 Annual Conference, Glasgow, Scotland, 3 February.

---. 2004. "A Comparison of Var and Cvar Constraints on Portfolio Selection with the Mean-Variance Model." *Management Science* 50(9): 1261-1273.

Alexander, G. J., A. M. Baptista, and S. Yan. 2007. "Mean-Variance Portfolio Selection with `at-Risk' Constraints and Discrete Distributions." *Journal of Banking & Finance* 31(12): 3761-3781.

Alexander, S., T. F. Coleman, and Y. Li. 2006. "Minimizing Cvar and Var for a Portfolio of Derivatives." *Journal of Banking & Finance* 30(2): 583-605.

Ane, T., and C. Kharoubi. 2003. "Dependence Structure and Risk Measure." *The Journal of Business* 76(3): 411-438.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath. 1999. "Coherent Measures of Risk." *Mathematical Finance* 9(3): 203.

Bai, M., and L. Sun. 2007. " Application of Copula and Copula-Cvar in the Multivariate Portfolio Optimization." In B. Chen, M. Paterson, and G. Zhang eds. *Combinatorics, Algorithms, Probabilistic and Experimental Methodologies*. Berlin, Germany: Springer, pp. 231-242.

Basak, G., R. Jagannathan, and G. Sun. 2002. "A Direct Test for the Mean Variance Efficiency of a Portfolio." *Journal of Economic Dynamics and Control* 26(7-8): 1195-1215.

Boehlje, M. D., and D. A. Lins. 1998. " Risks and Risk Management in an Industrialized Agriculture." *Agricultural Finance Review* 58(1): 1-16.

Bouyé, E., V. Durrleman, A. Nikeghbali, G. Riboulet, and T. Roncalli. 2001. "Copulas: An Open Field for Risk Management." Working paper, Warwick Business School, University of Warwick.

Buch, A., and G. Dorfleitner. 2008. "Coherent Risk Measures, Coherent Capital Allocations and the Gradient Allocation Principle." *Insurance: Mathematics and Economics* 42(1): 235-242.

Charpentier, A., and J. Segers. 2007. "Lower Tail Dependence for Archimedean Copulas: Characterization and Pitfalls." *Insurance: Mathematics and Economics* 40(3): 525-532.

Cherubini, U., E. Luciano, and W. Vecchiato. 2004. *Copula Methods in Finance*. Chichester, UK: John Wiley & Sons.

Clemen, R. T., and T. Reilly. 1999. "Correlations and Copulas for Decision and Risk Analsysis." *Management Science* 45(2): 208-224.

Cornuejols, G., and R. Tutuncu. 2009. *Optimization Methods in Finance*. Cambridge, MA: Cambridge University Press.

Crisostomo, M. F., and A. M. Featherstone. 1990. "A Portfolio Analysis of Returns to Farm Equity and Assets." *North Central Journal of Agricultural Economics* 12(1): 9-21.

Csóka, P., P. J.-J. Herings, and L. Á. Kóczy. 2007. "Coherent Measures of Risk from a General Equilibrium Perspective." *Journal of Banking & Finance* 31(8): 2517-2534.

Davis, S. B., T. J. Price, M. E. Wetzstein, and M. W. Rieger. 1997. "Reducing Yield Variation in Peach Orchards by Geographic Scattering." *American Journal of Agricultural Economics* 79(4): 1119-1126.

Dias, A. D. C. 2004. "Copula Inference for Finance and Insurance." Ph.D. dissertation, Swiss Federal Institute of Technology.

Dowd, K. 2005. "Copulas and Coherence." *Journal of Portfolio Management* 32(1): 123-127.

Duffie, D., and J. Pan. 1997. "An Overview of Value at Risk." *The Journal of Derivatives* 4(3): 7-49.

Edwards, C. 1959. "Resource Fixity and Farm Organization." *Journal of Farm Economics* 41(4): 747-759.

Ehling, P., and S. B. Ramos. 2006. "Geographic Versus Industry Diversification: Constraints Matter." *Journal of Empirical Finance* 13(4-5): 396-416.

Elton, E. J., and M. J. Gruber. 1997. "Modern Portfolio Theory, 1950 to Date." *Journal of Banking & Finance* 21(11-12): 1743-1759.

Embrechts, P., A. McNeil, and D. Strauman. 2002. "Correlation and Dependence in Risk Management: Properties and Pitfalls." In M. A. H. Dempster ed. *Risk Management: Value at Risk and Beyond*. Cambridge, MA: Cambridge University Press, pp. 176-223.

Escalante, C. L., and P. J. Barry. 2001. "Risk Balancing in an Integrated Farm Risk Management Plan." *Journal of Agricultural and Applied Economics* 33(3): 413-429.

Featherstone, A. M., P. V. Preckel, and T. G. Baker. 1990. "Modeling Farm Financial Decisions in a Dynamic and Stochastic Environment." *Agricultural Finance Review* 50(1): 80-99.

Featherstone, A. M., G. A. Ibendahl, J. R. Winter, and A. Spaulding. 2005. "Farm Financial Structure." *Agricultural Finance Review* 65(2): 97-117.

Flaten, O., and G. Lien. 2007. "Stochastic Utility-Efficient Programming of Organic Dairy Farms." *European Journal of Operational Research* 2007(181): 1574-1583.

Genest, C., and A. C. Favre. 2007. "Everything You Always Wanted to Know About Copula Modeling but Were Afraid to Ask." *Journal of Hydrologic Engineering* 2007(July/August): 347-368.

Goodwin, B. K., and A. P. Ker. 2002. "Modeling Price and Yield Risk." In R. E. Just, and R. D. Pope eds. *A Comprehensive Assessment of the Role of Risk in U.S. Agriculture*. Boston, MA: Kluwer Academic, pp. 289-323.

Gulpinar, N., B. Rustem, and R. Settergren. 2004. "Simulation and Optimization Approaches to Scenario Tree Generation." *Journal of Economic Dynamics & Control* 28(7): 1291-1315.

Handy, C., and J. M. MacDonald. 1989. "Multinational Structures and Strategies of U.S. Food Firms." *American Journal of Agricultural Economics* 71(5): 1246-1254.

Hardaker, J. B., S. Pandey, and L. H. Patten. 1991. "Farm Planning under Uncertainty: A Review of Alternative Programming Models." *Review of Marketing and Agricultural Economics* 59(1): 9-22.

Hardaker, J. B., R. B. M. Huirne, J. R. Anderson, and G. Lien. 1997. *Coping with Risk in Agriculture*. 2nd. ed. Cambridge, MA: CABI.

Harwood, J., R. G. Heifner, J. Perry, and A. Somwaru. 1999. *Managing Risk in Farming: Concepts, Research and Analysis*. Washington DC: United States Department of Agriculture E. R. Service, Rep. 774, Heady, E. O. 1952. "Diversification in Resource Allocation and Minimization of Income Variability." *Journal of Farm Economics* 34(4): 482-496.

Hennessy, D. A., and H. E. Lapan. 2002. "The Use of Archimedean Copulas to Model Portfolio Allocations." *Mathematical Finance* 12(2): 143-154.

Hennings, E., B. J. Sherrick, and P. J. Barry. 2005. "Portfolio Diversification Using Farmland Investments." Paper presented at American Agricultural Economics Association Annual Meeting, Providence, Rhode Island, 24-27 July.

Hochreiter, R. 2010. "Evolutionary Multi-Stage Financial Scenario Tree Generation." In C. Di Chio, et al. eds. *Applications of Evolutionary Computation*. Berlin, Germany: Springer pp. 182-191.

Hoyland, K., and S. W. Wallace. 2001. "Generating Scenario Trees for Multistage Decision Problems." *Management Science* 47(2): 295-307.

Infanger, G. 2006. "Dynamic Asset Allocation Strategies Using a Stochastic Dynamic Programming Approach." In S. A. Zenios, and W. T. Ziemba eds. *Handbook of Asset and Liability Management*. North Holland, The Netherlands: Elsevier, pp. 200-248.

Joe, H. 1997. *Multivariate Models and Dependence Concepts*. London, UK: Chapman & Hall.

Johnson, M. A., and E. C. J. Pasour. 1981. "An Opportunity Cost View of Fixed Asset Theory and the Overporduction Trap." *American Journal of Agricultural Economics* 63(1): 1-7.

Jorion, P. 1996. "Risk2: Measuring the Risk in Value at Risk." *Financial Analysts Journal* Nov/Dec 1996(52): 47-56.

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

Just, R. E., and R. D. Pope. 2002. "Past Progress and Future Opportunities for Agricultural Risk Research." In R. E. Just, and R. D. Pope eds. *A Comprehensive Assessment of the Role of Risk in U.S. Agriculture*. Boston, MA: Kluwer Academic, pp. 557-578.

---. 2003. "Agricultural Risk Analysis: Adequacy of Models, Data, and Issues." *American Journal of Agricultural Economics* 85(5): 1249-1256.

Kaiser, E., and M. D. Boehlje. 1980. "A Multiperiod Risk Programming Model for Farm Planning." *North Central Journal of Agricultural Economics* 2(1): 47-54.

Kaiser, H. M., and J. Appland. 1989. "Dssp: A Model of Production and Marketing Decisions on a Midwestern Crop Farm." *North Central Journal of Agricultural Economics* 11(2): 157-169.

Kallenberg, W. C. M. 2008. "Modeling Dependence." *Insurance: Mathematics and Economics* 42(1): 127-146.

Kaut, M., and S. W. Wallace. 2007. "Evaluation of Scenario-Generation Methods for Stochastic Programming." *Pacific Journal of Optimization* 3(2): 257-271.

---. 2009. "Shape-Based Scenario Generation Using Copulas." *Computational Management Science* 8(1-2): 181-199.

Kim, Y. S., and I. Mathur. 2008. "The Impact of Geographic Diversification on Firm Performance." *International Review of Financial Analysis* 2008(17): 747-766.

Krokhmal, P. A. 2007. "Higher Moment Coherent Risk Measures." *Quantitative Finance* 7(4): 373-387.

Krueger, A., V. Salin, G. Williams, L. Eden, and A. Gray. 1999. "Profitability of Geographic Diversification Strategy." *Journal of Food Distribution Research* 30(1): 112-123.

Krueger, A., V. Salin, and A. W. Gray. 2002. "Geographic Diversification Strategy and the Implications of Global Market Integration in Table Grapes." *Agribusiness* 18(1): 81-99.

Lambert, D. K., and B. A. McCarl. 1989. "Sequential Modeling of White Wheat Marketing Strategies." *North Central Journal of Agricultural Economics* 11(1): 105-115.

Larsen, R. Personal Interviews at 2008 Apex Annual Meetings, San Juan, Puerto Rico.

Leatham, D. J., and T. G. Baker. 1988. "Farmer's Choice of Fixed and Adjustable Interest Rate Loans." *American Journal of Agricultural Economics* 70(4): 803-812.

Liang, N., and S. A. Rhoades. 1988. "Geographic Diversification and Risk in Banking." *Journal of Economics and Business* 40(4): 271-284.

Lien, G., J. Hardaker, M. van Asseldonk, and J. Richardson. 2009. "Risk Programming Analysis with Imperfect Information." *Annals of Operations Research* 1-13.

Manfredo, M. R., and R. M. Leuthold. 2001. "Market Risk and the Cattle Feeding Margin: An Application of Value-at-Risk." *Agribusiness* 17(3): 333-353.

Markowitz, H. 1952. "Portfolio Selection." *The Journal of Finance* 7(1): 77-91.

McCarl, B. A., and D. A. Bessler. 1989. "Estimating an Upper Bound on the Pratt Risk Aversion Coefficient When the Utility Function Is Unknown." *Australian Journal of Agricultural Economics* 33(1): 56-63.

McNeil, A., R. Frey, and P. Embrechts 2005. *Quantitative Risk Management.* D. Duffie, ed. Princeton, NJ: Princeton University Press.

Mishra, A., and H. El-Osta. 2002. "Risk Management through Enterprise Diversification: A Farm-Level Analysis." Paper presented at American Agricultural Economics Association Annual Meeting, Long Beach, CA, 28-31 July.

Mishra, A., H. El-Osta, and C. L. Sandretto. 2004. "Factors Affecting Farm Enterprise Diversification." *Agricultural Finance Review* 64 (2): 151-166.

Mishra, A., and S. H. Lence. 2005. "Risk Management by Farmers, Agribusinesses, and Lenders." *Agricultural Finance Review* 65(2): 131-148.

Moschini, G., and D. A. Hennessy. 2001. "Uncertainty, Risk Aversion and Risk Management for Agricultural Producers." In B. L. Gardner, and G. C. Rausser eds. *Handbook of Agricultural Economics*. New York, NY: Elsevier, pp. 87-144.

Mulvey, J. M., and B. Shetty. 2004. "Financial Planning Via Multi-Stage Stochastic Optimization." *Computers & Operations Research* 31(1): 1-20.

Nartea, G. V., and P. J. Barry. 1994. "Risk Efficiency and Cost Effects of Geographic Diversification." *Review of Agricultural Economics* 16(3): 341-351.

Nelsen, R. B. 2006. *An Introduction to Copulas*. 2nd ed. New York, NY: Springer.

Ozaki, V. A., B. K. Goodwin, and R. Shirota. 2008. "Parametric and Nonparametric Statistical Modelling of Crop Yield: Implications for Pricing Crop Insurance Contracts." *Applied Economics* 40(9): 1151-1164.

Patton, A. J. 2002. "Applications of Copula Theory in Financial Econometrics." Ph.D. dissertation, University of California, San Diego.

Peterson, H. H., B. Schurle, and M. Langemeier. 2005. "Three-Dimensional Efficient Portfolio Frontier: Mean, Variance, and Farm Size." Paper presented at American Agricultural Economics Association Annual Meeting, Providence, Rhode Island, 24-27 July.

Pope, R. D., and R. Prescott. 1980. "Diversification in Relation to Farm Size and Other Socioeconomic Characteristics." *American Journal of Agricultural Economics* 62(3): 554-559.

Power, G. J., and D. Vedenov. 2008. "The Shape of the Optimal Hedge Ratio: Modeling Joint Spot-Futures Prices Using an Empirical Copula-Garch Model " Paper presented at NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management, St. Louis, MO, 21-22 April.

Rae, A. N. 1971a. "Decision Problems in Farm Management." *American Journal of Agricultural Economics* 53(3): 448-460.

---. 1971b. "An Empirical Application and Evaluation of Discrete Stochastic Programming in Farm Management." *American Journal of Agricultural Economics* 53(4): 625-638.

Rank, J. 2000. "Copulas in Financial Risk Management." Ph.D. dissertation, University of Oxford.

Robison, L. J., and J. R. Brake. 1979. "Application of Portfolio Theory to Farmer and Lender Behavior." *American Journal of Agricultural Economics* 61(1): 158-164.

Robison, L. J., and P. J. Barry. 1980. "Portfolio Theory and Asset Indivisibility: Implications for Analysis of Risk Management." *North Central Journal of Agricultural Economics* 2(1): 41-46.

Rockafellar, R. T., and S. Uryasev. 2000. "Optimization of Conditional Value-at-Risk." *Journal of Risk* 2(3): 21-41.

Rose, J. T., and J. D. Wolken. 1990. "Geographic Diversification in Banking, Market Share Changes, and Viability of Small Independent Banks." *Journal of Financial Services Research* 4(1): 5-20.

Schmidt, T. 2006. "Coping with Copulas." In J. Rank ed. *Copulas, from Theory to Application in Finance*. London, UK: Incisive Financial, pp. 3-34.

Sharpe, W. F. 1963. "A Simplified Model for Portfolio Analysis." *Management Science* 9(2): 277-293.

Sklar, A. 1959. "Fonction De Repartiton a N Dimensions Et Leurs Marges." *Publications de I'Institut de Statistique de L'Universite de Paris* 8(1): 229-231.

Stoica, G. 2006. "Relevant Coherent Measures of Risk." *Journal of Mathematical Economics* 42(6): 794-806.

Sun, W., S. Rachev, F. Fabozzi, and P. Kalev. 2009. "A New Approach to Modeling Co-Movement of International Equity Markets: Evidence of Unconditional Copula-Based Simulation of Tail Dependence." *Empirical Economics* 36(1): 201-229.

Szegö, G. 2005. "Measures of Risk." *European Journal of Operational Research* 163(1): 5-19.

Tobin, J. 1958. "Liquidity Preference as Behavior Towards Risk." *The Review of Economic Studies* 25(2): 65-86.

Topaloglou, N., H. Vladimirou, and S. A. Zenios. 2008a. "Pricing Options on Scenario Trees." *Journal of Banking & Finance* 32(2): 283-298.

---. 2008b. "A Dynamic Stochastic Programming Model for International Portfolio Management." *European Journal of Operational Research* 185(3): 1501-1524.

Torkamani, J., and J. B. Hardaker. 1996. "A Study of Economic Efficiency of Iranian Farmers in Ramjerd District: An Application of Stochastic Programming." *Agricultural Economics* 14(2): 73-83.

Torkamani, J. 2005. "Using a Whole-Farm Modelling Approach to Assess Prospective Technologies under Uncertainty." *Agricultural Systems* 85(2): 138-154.

Trivedi, P. K., and D. M. Zimmer. 2005. "Copula Modeling: An Introduction for Practitioners." *Foundations and Trends in Econometrics* 1(1): 1-111.

Turvey, C. G., H. C. Driver, and T. G. Baker. 1988. "Systematic and Nonsystematic Risk in Farm Portfolio Selection." *American Journal of Agricultural Economics* 70(4): 831-836.

USDA-NASS. National Agriculture Statistics Service/State Facts/County Facts Internet Site: www.nass.usda.gov/QuickStats/ (Accessed March 2008).

USDA-NRCS. Natural Resource and Conservation Service Internet Site: www.wcc.nrcs.usda.gov/climate/wetlands.html (Accessed March 2008).

USDA. 2000. *Off Farm Income Supports Many Farm Households*. E. R. S. Food and Rural Economics Division, Rep. AIS-74, Washington DC, February.

USDC. Bureau of Economic Analysis, National Economic Accounts, Internet Site: www.bea.gov, (Accessed November 2009).

Vedenov, D. V., and B. J. Barnett. 2004. "Efficiency of Weather Derivatives as Primary Crop Insurance Instruments." *Journal of Agricultural and Resource Economics* 29(3): 387-403.

Vedenov, D. V. 2008. "Application of Copulas to Estimation of Joint Crop Yield Distributions." Paper presented at American Agricultural Economics Association Annual Meeting, Orlando, FL, 27-29 July.

Wackerly, D. D., W. Mendenhall III, and R. L. Scheaffer. 2002. *Mathematical Statistics with Applications*. 6th ed. Pacific Grove, CA: Duxbury.

Wolfley, J. L. 2008. "Machinery Sharing by Agribusiness Firms: Methodology, Application, and Simulation." Ph.D. dissertation, Texas A&M University.

Xu, Y. 2005. "Applications of Copula-Based Models in Portfolio Optimization." Ph.D. dissertation, University of Miami.

Zhu, Y., S. K. Ghosh, and B. K. Goodwin. 2008. ""Modeling Dependence in the Design of Whole Farm---a Copula-Based Model Approach." Paper presented at American Agricultural Economics Association Annual Meeting, Orlando, FL, 27-29 July.

Zylstra, M. J., R. L. Kilmer, and S. Urysasev. 2003. "An Application of Conditonal Value at Risk." Paper presented at American Agricultural Economics Association Meetings, Montreal, Canada, 27-30 July.

APPENDIX A

Table A.1. Trend Regression Results

State	Total[1]	Trend	%	Total	Trend	%	Total	Trend	%
	Wheat			Cotton			Sorghum		
Colorado	18	17	94				1	1	100
Idaho	10	10	100						
Montana	32	22	69						
Nebraska	37	29	78				7	3	43
New Mexico							2	0	0
North Dakota	45	45	100						
Oregon	11	9	82						
South Dakota	37	25	68				2	2	100
Utah	4	0	0						
Washington	8	3	38						
Kansas	104	5	5				23	18	78
Oklahoma	8	1	13	1	0	0	2	1	50
Texas	66	12	18	20	2	10	13	5	38
Arkansas				3	2	67			
Louisiana				7	0	0			
Mississippi				8	0	0			
Total	380	178	47	39	4	10	50	30	60

1) Total is the total number of counties used in the analysis, trend is the number of counties with a significant trend, and % is the percent with a significant trend.

Significance at the 0.05 level is used for trend determination.

Table A.2. Variable Descriptive Statistics

Absolute Value of Change in	Mean	Standard Deviation	Minimum	Maximum
	Wheat - number of observations 72,010			
Elevation (ft)	1330.23	1064.82	0.00	6916
Precipitation (in)	9.499926	7.666816	0.00	49.73
Temperature (°f)	7.958991	5.843862	0.00	31.70
Correlation	0.164671	0.292902	-0.74	0.97
	Cotton - number of observations 741			
Elevation (ft)	1450.34	1165.20	0.00	3710.00
Precipitation (in)	18.92	14.57	0.02	44.02
Temperature (°f)	2.68	1.94	0.00	9.89
Correlation	0.33	0.28	-0.48	0.94
	Sorghum - number of observations 1,225			
Elevation (ft)	1424.34	1078.04	0.00	4380.00
Precipitation (in)	8.47	5.98	0.02	28.50
Temperature (°f)	6.57	6.41	0.00	25.80
Correlation	0.31	0.29	-0.46	0.97

Table A.3. Copula Characteristics

Normal Copula	
CDF	$C(u,v;\rho) = \theta_\rho(\theta^{-1}(u), \theta^{-1}(v))$
PDF	$c(u,v) = \frac{1}{\sqrt{1-\rho^2}} exp\left(\frac{\left(\Phi^{-1}(u)\right)^2 + \left(\Phi^{-1}(v)\right)^2}{2} \frac{2\rho\Phi^{-1}(u)\Phi^{-1}(v) - \left(\Phi^{-1}(u)\right)^2 - \left(\Phi^{-1}(v)\right)^2}{2(1-\rho^2)}\right)$
Parameter Range	$\rho \in (-1,1)$
Kendall's tau	$\tau_\rho = \frac{2\arcsin(\rho)}{\pi}$
Tail Dependency	$\lambda_L = \lambda_U = 0$
Clayton Copula	
CDF	$C(u,v;\theta) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}$
PDF	$c(u,v;\theta) = (1+\theta)(uv)^{-\theta-1}\left(u^{-\theta} + v^{-\theta} - 1\right)^{-2-\frac{1}{\theta}}$
Parameter Range	$\theta \in [-1,\infty)\backslash\{0\}$
Kendall's tau	$\tau_\theta = \frac{\theta}{\theta+2}$
Tail Dependency	$\lambda_L = 2^{-\frac{1}{\theta}}, \lambda_U = 0$
Gumbel Copula	
CDF	$C(u,v{:}\,\delta) = exp\left\{-\left((-\log u)^\delta + (-\log v)^\delta\right)^{\frac{1}{\delta}}\right\}$
PDF	$c(u,v{:}\,\delta) = \frac{C(u,v;\delta)(\log u \cdot \log v)^{\delta-1}}{uv((-\log u)^\delta + (-\log v)^\delta)^{2-\frac{1}{\delta}}}\left(((-\log u)^\delta + (-\log v)^\delta)^{\frac{1}{\delta}} + \delta - 1\right)$
Parameter Range	$\delta \in [1,\infty)$
Kendall's tau	$\tau_\delta = 1 - \frac{1}{\delta}$
Tail Dependency	$\lambda_L = 0, \lambda_U = 2 - 2^{1/\delta}$
Rotated Gumbel Copula	
CDF	$C(u,v{:}\,\delta) = u + v - 1 + C(1-u, 1-v;\delta)$
PDF	$c(u,v{:}\,\delta) = c(1-u, 1-v;\delta)$
Parameter Range	$\delta \in [1,\infty)$
Kendall's tau	$\tau_\delta = 1 - \frac{1}{\delta}$
Tail Dependency	$\lambda_L = 2 - 2^{1/\delta}, \lambda_U = 0$

Table A.4. Model Comparison with Linear Correlation as Dependency Measure

Model	Degrees of Freedom	AIC[1]	BIC[2]	R^2
		Wheat		
Linear	4	10109.10	10145.84	0.22
Log-Linear	4	9382.86	9419.59	0.22
Quadratic	9	-44.44	38.22	0.32
		Cotton		
Linear	4	-7.72	10.71	0.25
Log-Linear	4	-13.13	5.25	0.25
Quadratic	9	-92.27	-46.20	0.34
		Sorghum		
Linear	4	62.92	83.36	0.29
Log-Linear	4	81.80	102.20	0.27
Quadratic	9	-137.08	-91.08	0.40

1) Akaike Information Criterion defined as AIC = $e^{\left(\frac{2k}{T}\right)}MSE$.

Bayesian Information Criterion defined as BIC = $T^{\left(\frac{k}{T}\right)}MSE$.

Table A.5. Linear Regression Results for Non-Irrigated and Linear Correlation

Variable	Coefficient	t-value	p-value
	Wheat		
Elevation	-0.00001	-13.66	0.00
Temperature	-0.01338	-80.87	0.00
Precipitation	-0.01171	-94.30	0.00
Intercept	0.40077	170.06	0.00
$R^2 = 0.22$			
	Cotton		
Elevation	0.0001	5.46	0.00
Temperature	-0.0048	-7.27	0.00
Precipitation	-0.0123	-12.10	0.00
Intercept	0.4949	27.98	0.00
$R^2 = 0.25$			
	Sorghum		
Elevation	-0.00005	-5.45	0.00
Temperature	-0.01962	-15.49	0.00
Precipitation	-0.00642	-4.35	0.00
Intercept	0.55635	35.61	0.00
$R^2 = 0.29$			

Table A.6. Log-Linear Regression Results for Non-Irrigated and Linear Correlation

Variable	Coefficient	t-value	p-value
	Wheat		
Elevation	-0.02	-27.32	0.00
Temperature	-0.09	-90.55	0.00
Precipitation	-0.07	-73.41	0.00
Intercept	0.59	98.02	0.00
$R^2 = 0.22$			
	Cotton		
Elevation	0.04	4.65	0.00
Temperature	-0.05	-6.78	0.00
Precipitation	-0.11	-10.05	0.00
Intercept	0.38	8.62	0.00
$R^2 = 0.25$			
	Sorghum		
Elevation	-0.08	-9.11	0.00
Temperature	-0.08	-12.91	0.00
Precipitation	-0.01	-1.57	0.12
Intercept	0.94	18.68	0.00
$R^2 = 0.27$			

Table A.7. Quadratic Regression Results for Linear Correlation

Variable	Wheat	Cotton	Sorghum
Elevation	-0.00011 (-39.21)**	0.0001 (3.63)**	-0.00025 (-5.11)**
Temperature	-0.05072 (-95.16)**	-0.0363 (-2.91)**	-0.06469 (-14.30)**
Precipitation	-0.02371 (-56.40)**	-0.0358 (-10.77)**	-0.00251 (-0.69)
$Elevation^2$	-1.07E09 (-1.90)	-7.96E-08 (-4.21)**	4.18E-08 (1.43)
$Temperature^2$	0.00112 (52.73)**	-0.0055 (2.65)**	0.00152 (7.15)**
$Precipitation^2$	-0.00013 (-8.42)**	0.0004 (2.51)**	-1E-05 (-0.05)
Elev*Temp	0.000006 (33.26)**	0.00003 (4.70)**	1.04E-05 (-8.41)88
Elev*Precip	0.000005 (43.61)**	5.11E-06 (2.16)**	8.50E-07 (0.46)
Precip*Temp	0.000744 (33.55)**	-0.0001 (-0.17)	-3.41E-12 (-0.77)
Intercept	0.646415 (185.33)	0.6407 (23.21)	0.75552 (29.08)
R^2	0.32	0.33	0.40

Notes: Single and double asterisks (*) denote statistical significance at the 0.05 and 0.01 levels. Numbers in parentheses are calculated t-statistic values.

Table A.8. Elasticities (estimated at mean) for Linear Correlation

Variable	Wheat	Cotton	Sorghum
	Linear Model		
Elevation	0.11	0.31	0.22
Temperature	0.65	0.11	0.42
Precipitation	0.68	0.71	0.18
	Log-Linear Model		
Elevation	0.27	0.15	0.35
Temperature	0.99	0.18	0.37
Precipitation	0.75	0.43	0.06
	Quadratic Model		
Elevation	0.18	0.76	0.42
Temperature	1.28	0.19	1.40
Precipitation	1.09	1.23	0.06

Table A.9.   Wheat Quadratic Regression Coefficients

Variable	Linear Correlation	Gaussian Copula	Frank Copula	Clayton Copula	Gumbel Copula	Clayton Tail Dependence	Gumbel Tail Dependence
Elevation	-0.00011 (-39.21)**	-0.0001 (-36.11)**	-6.1E-05 (-36.74)**	-6.5E-05 (-42.52)**	-5.9E-05 (-34.44)**	-9.3E-05 (-40.92)**	-7.5E-05 (-38.43)**
Temperature	-0.05072 (-95.16)**	-0.04998 (-91.47)**	-0.03123 (-96.4)**	-0.02767 (-93.88)**	-0.03227 (-96.17)**	-0.04101 (-95.69)**	-0.03667 (-97.68)**
Precipitation	-0.02371 (-56.40)**	-0.02233 (-54.92)**	-0.01506 (-62.85)**	-0.01462 (-63.82)**	-0.01434 (-58.36)**	-0.02191 (-65.53)**	-0.01822 (-62.61)**
Elevation^2	-1.07E09 (-1.90)	-3.57E-10 (1.26)	-1.50E-10 (-0.47)	4.44E-10 (1.5)	-6.42E-10 (-1.91)	-2.25E-10 (-0.51)	-5.54E-10 (-1.43)
Temperature^2	0.00112 (52.73)**	0.001128 (54.33)**	0.000625 (51.54)**	0.000492 (43.82)**	0.00066 (51.09)**	0.000683 (41.78)**	0.000677 (46.12)**
Precipitation^2	-0.00013 (-8.42)**	-0.00014 (-5.44)**	-1.5E-05 (-1.97)*	-2.5E-05 (-3.42)**	-3.6E-05 (-4.69)**	-2.7E-05 (-2.47)**	-5.4E-05 (-5.61)**
Elev*Temp	0.000006 (33.26)**	5.23E-06 (30.53)**	3.12E-06 (33.48)**	2.71E-06 (31.5)**	3.56E-06 (36.32)**	4.13E-06 (32.80)**	3.65E-06 (32.24)**
Elev*Precip	0.000005 (43.61)**	5.09E-06 (39.29)**	2.77E-06 (45.17)**	2.91E-06 (49.23)**	2.68E-06 (41.6)**	4.24E-06 (49.94)**	3.62E-06 (46.71)**
Precip*Temp	0.000744 (33.55)**	0.0007 (31.63)**	0.000571 (49.6)**	0.000575 (52.59)**	0.000559 (45.59)**	0.000896 (57.35)**	0.00074 (51.32)**
Intercept	0.646415	0.622777	0.44509	0.431719	0.45408	0.608691	0.558589

Notes: Single and double asterisks (*) denote statistical significance at the 0.05 and 0.01 levels. Numbers in parentheses are calculated t-statistic values.

Table A.10. Estimated Absolute Value of the Elasticities Evaluated at the Mean

Variable	Linear Correlation	Gaussian Copula	Frank Copula	Clayton Copula	Gumbel Copula	Clayton Tail Dependence	Gumbel Tail Dependence
				Wheat			
Elevation	0.18	0.13	0.11	0.15	0.07	0.17	0.10
Temperature	1.28	1.33	0.77	0.69	0.70	0.82	0.65
Precipitation	1.09	1.09	0.56	0.51	0.50	0.59	0.47
				Cotton			
Elevation	1.92	1.27	1.05	0.78	0.89	0.81	0.80
Temperature	3.95	2.34	2.06	1.19	1.47	1.25	1.35
Precipitation	1.23	0.75	0.67	0.57	0.62	0.56	0.55
				Sorghum			
Elevation	0.61	3.17	1.30	0.99	1.09	1.24	0.98
Temperature	1.02	3.47	1.10	1.02	1.03	1.12	0.96
Precipitation	0.09	0.58	0.18	0.10	0.12	0.13	0.15

Table A.11. Sorghum Quadratic Regression Coefficients

Variable	Linear Correlation	Gaussian Copula	Frank Copula	Clayton Copula	Gumbel Copula	Clayton Tail Dependence	Gumbel Tail Dependence
Elevation	-0.00025 (-5.11)**	-0.00022 (-11.16)**	-0.00022 (-13.21)**	-0.00016 (-10.81)**	-0.0002 (-11.96)**	-0.00025 (-11.58)**	-0.00022 (-11.60)**
Temperature	-0.06469 (-14.30)**	-0.05383 (-13.68)**	-0.04413 (-14.51)**	-0.03922 (-14.19)**	-0.04405 (-14.62)**	-0.05398 (-13.87)**	-0.05085 (-14.67)**
Precipitation	-0.00251 (-0.69)	0.004698 (1.22)	0.000406 (0.12)	-0.0025 (-0.87)	-0.00175 (-0.55)	-0.00264 (-0.62)	-0.00062 (-0.17)
$Elevation^2$	4.18E-08 (1.43)	1.73E-08 (2.5)**	1.18E-08 (2.41)**	8.80E-09 (1.99)*	9.71E-09 (2.02)*	1.32E-08 (2.06)*	1.03E-08 (1.77)
$Temperature^2$	0.00152 (7.15)**	0.001407 (8.35)**	0.001105 (9.1)**	0.001026 (9.18)**	0.001123 (9.28)**	0.001368 (8.62)**	0.001305 (9.06)**
$Precipitation^2$	-1E-05 (-0.05)	-8.6E-05 (-0.48)	-8.6E-05 (-0.6)	3.58E-05 (0.28)	-2.2E-05 (-0.16)	-1.2E-05 (-0.06)	-1.9E-05 (-0.12)
Elev*Temp	1.04E-05 (-8.41)88	8.28E-06 (7.09)**	8.39E-06 (9.7)**	6.08E-06 (7.69)**	7.54E-06 (8.85)**	9.36E-06 (8.04)**	8.61E-06 (8.37)**
Elev*Precip	8.50E-07 (0.46)	1.35E-06 (0.8)	3.16E-06 (2.62)**	2.25E-06 (2.1)*	2.95E-06 (2.54)**	3.86E-06 (2.42)*	3.03E-06 (2.17)*
Precip*Temp	-3.41E-12 (-0.77)	-0.00028 (-1.32)	-0.00016 (-1.01)	-3.61E-06 (-0.03)	-5.5E-05 (0.35)	-8.3E-05 (-0.38)	-0.0001 (-0.56)
Intercept	0.75552	0.514015	0.539399	0.469676	0.533657	0.657855	0.615689

Notes: Single and double asterisks (*) denote statistical significance at the 0.05 and 0.01 levels. Numbers in parentheses are calculated t-statistic values.

Table A.12. Cotton Quadratic Regression Coefficients

Variable	Linear Correlation	Gaussian Copula	Frank Copula	Clayton Copula	Gumbel Copula	Clayton Tail Dependence	Gumbel Tail Dependence
Elevation	0.0001 (3.63)**	9.44E-05 (3.68)**	0.0001 (3.70)**	0.0001 (1.93)*	7.98E-05 (3.03)**	-0.0003 (-7.23)**	9.04E-05 (2.99)**
Temperature	-0.0363 (-2.91)**	-0.0315 (-3.46)**	-0.0318 (-3.47)**	-0.0206 (-2.32)*	-0.0254 (-2.65)**	-0.0063 (-0.44)	-0.0257 (-2.35)*
Precipitation	-0.0358 (-10.77)**	-0.0213 (-9.46)**	-0.0215 (-9.63)**	-0.0148 (7.10)**	-0.0196 (-8.70)**	0.0049 (1.32)	-0.0212 (-8.16)**
$Elevation^2$	-7.96E-08 (-4.21)**	-5.77E-08 (-4.50)**	-6.13E-08 (-4.95)**	-5.99E-08 (-4.77)**	-6.47E-08 (-4.93)**	3.00E-08 (1.54)	-7.59E-08 (-4.91)**
$Temperature^2$	-0.0055 (2.65)**	-0.0023 (-1.58)	-0.0024 (-1.69)	-0.0018 (-1.28)	-0.0024 (-1.55)	0.0040 (1.69)*	-0.0029 (-1.63)
$Precipitation^2$	0.0004 (2.51)**	0.0001 (1.29)	0.0001 (1.28)	-5.2E-05 (-0.54)	5.81E-05 (0.57)	-0.0003 (-1.82)*	2.15E-05 (0.18)
Elev*Temp	0.00003 (4.70)**	1.93E-05 (3.92)**	1.97E-05 (4.11)**	2.05E-05 (4.08)**	0.00002 (4.31)**	-9.02E-07 (-0.12)	2.51E-05 (4.22)**
Elev*Precip	5.11E-06 (2.16)**	4.78E-06 (2.94)**	5.03E-06 (3.26)**	7.30E-06 (4.59)**	6.38E-06 (3.86)**	6.22E-06 (2.49)**	7.81E-06 (4.02)**
Precip*Temp	-0.0001 (-0.17)	4.19E-05 (0.11)	0.00004 (0.11)	-0.0007 (-2.00)*	-0.0005 (-1.28)	-0.0006 (-0.96)	-0.0006 (-1.40)
Intercept	0.6407	0.4346	0.4844	0.4269	0.4862	0.5880	0.5597

Notes: Single and double asterisks (*) denote statistical significance at the 0.10 and 0.05 levels. Numbers in parentheses are calculated t-statistic values.

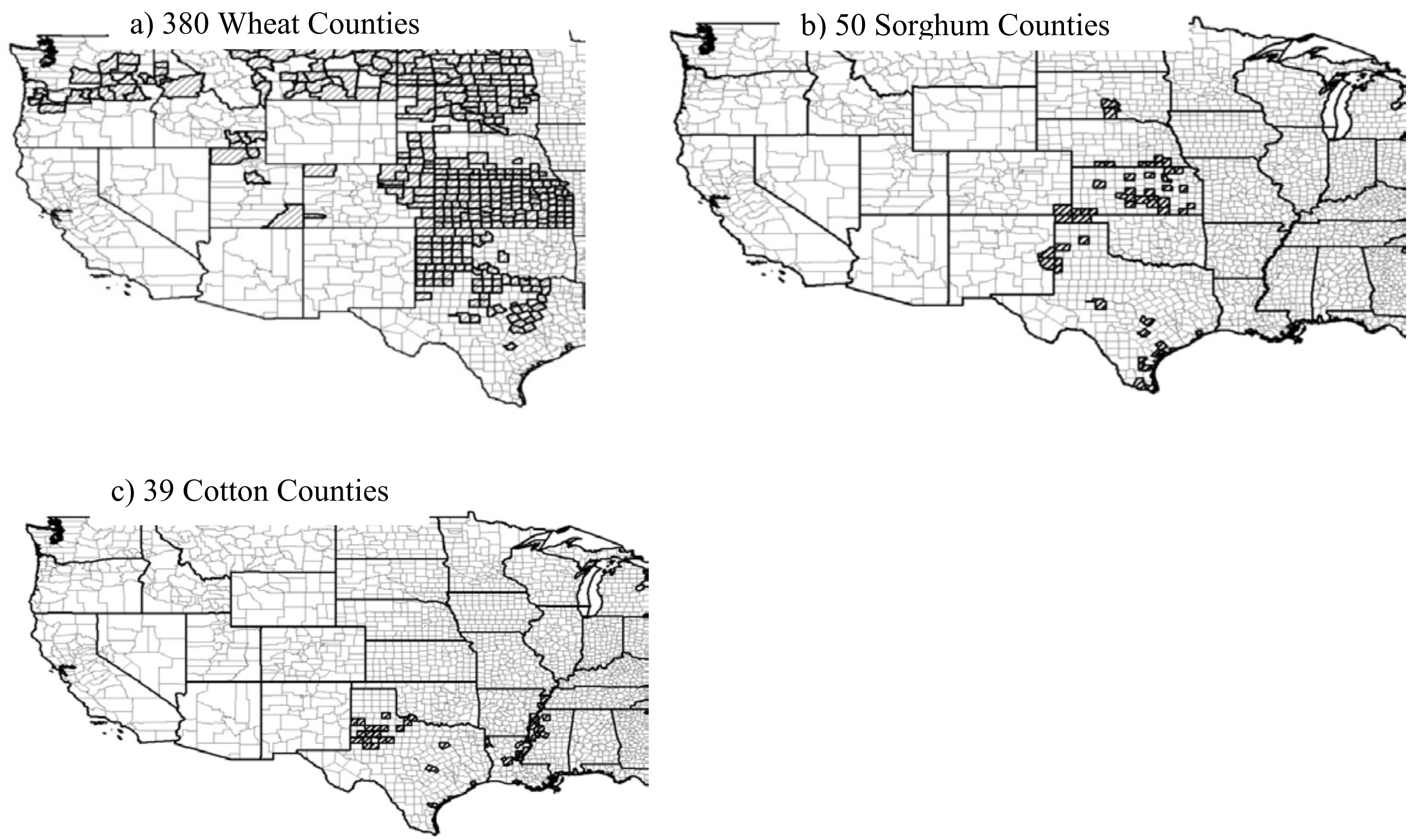

Figure A.1. Wheat, sorghum, and cotton counties included in analysis

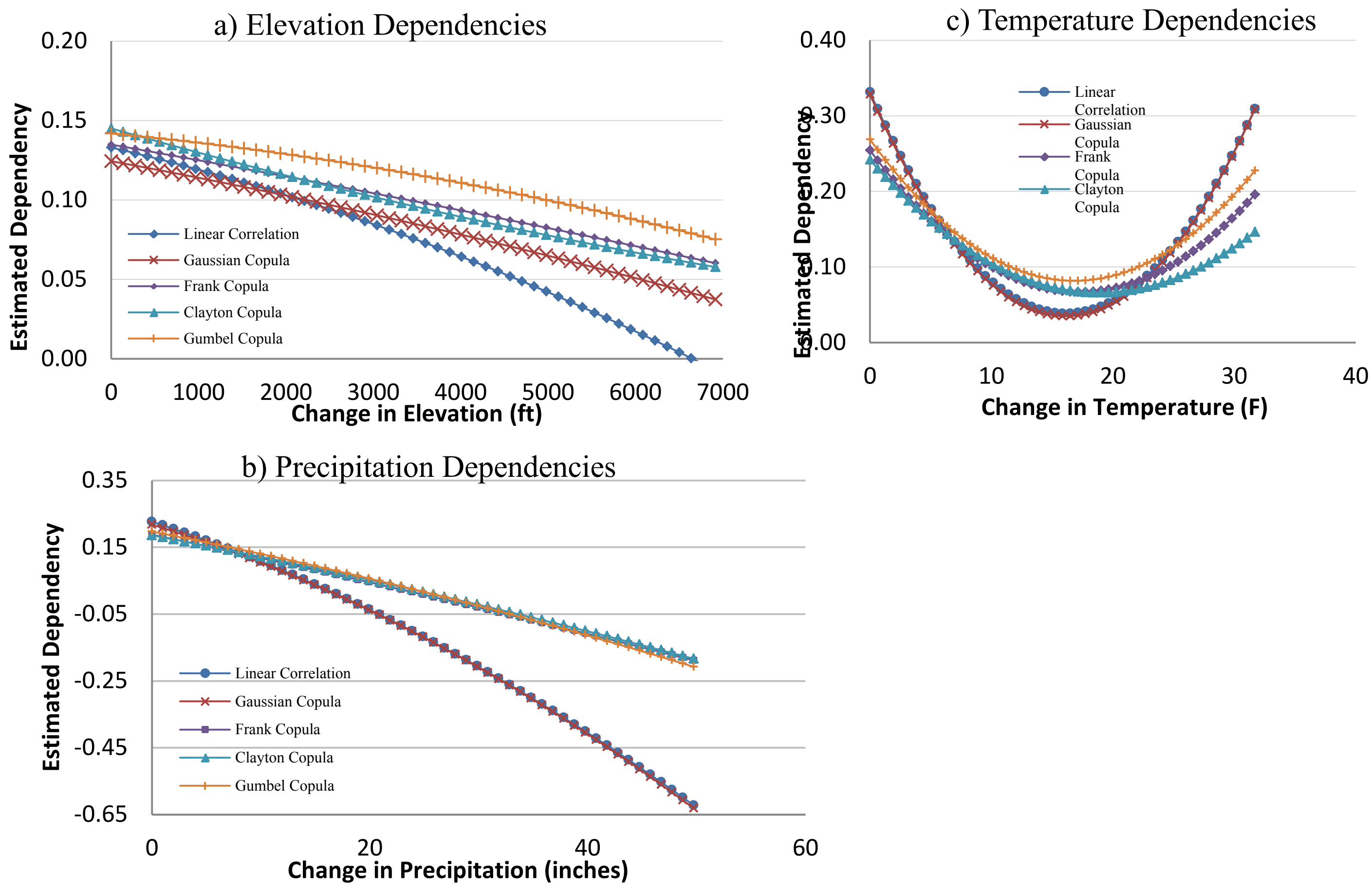

Figure A.2. Estimated wheat dependencies for elevation, precipitation, and temperature holding two variables constant while varying the three variables given by the x-axis

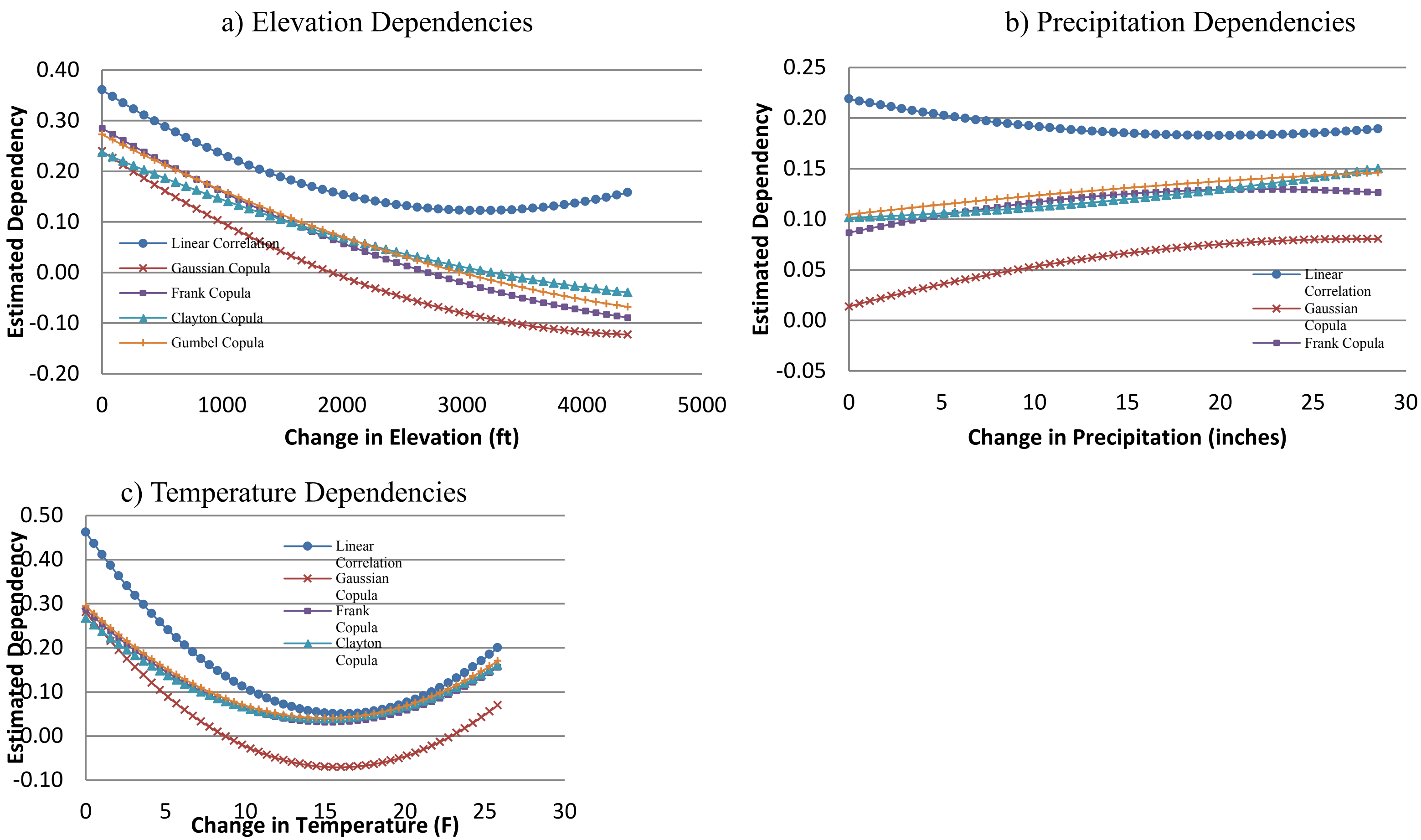

Figure A.3. Estimated sorghum dependencies for elevation, precipitation, and temperature holding two variables constant while varying the three variables given by the x-axis

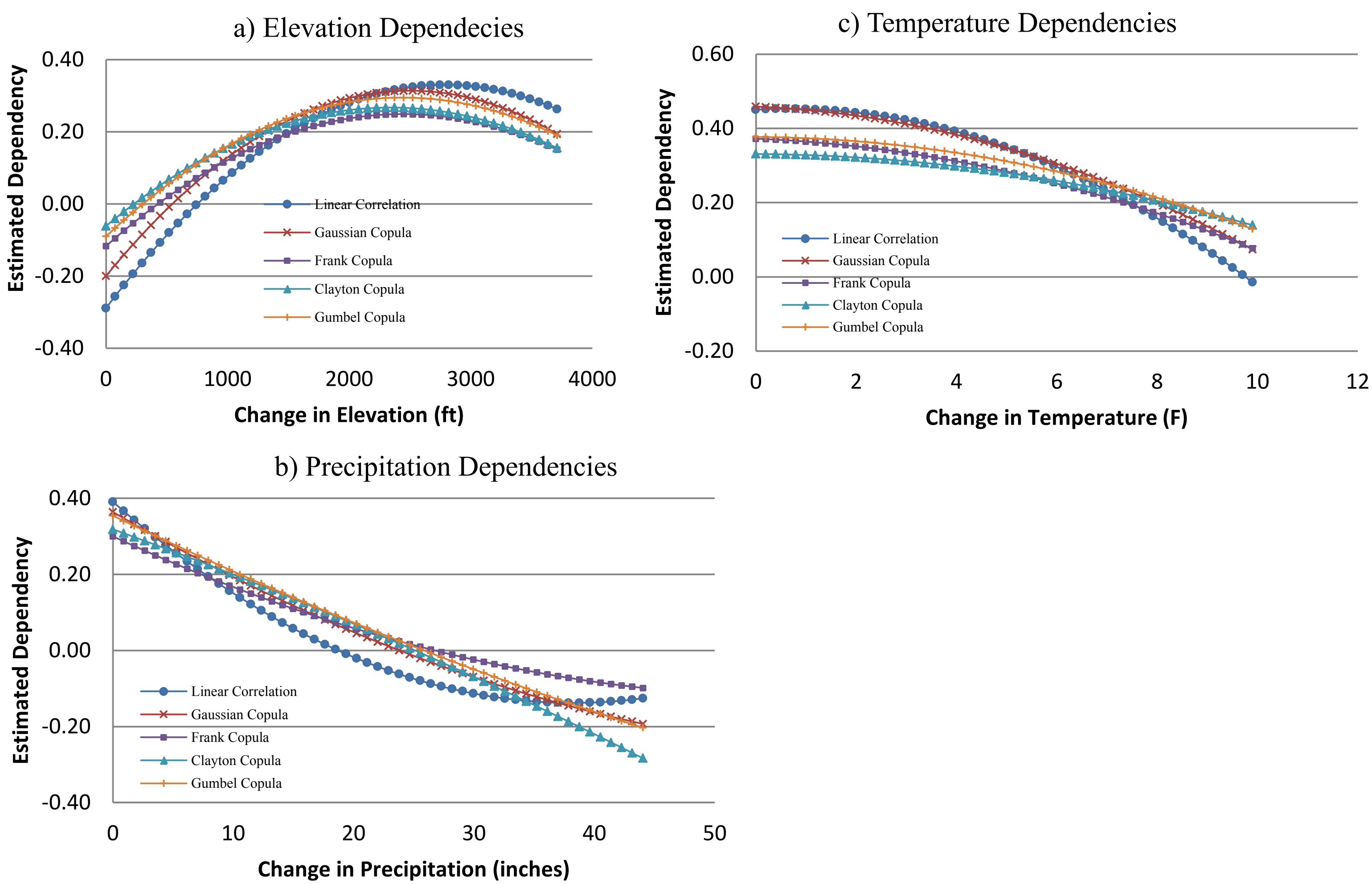

Figure A.4. Estimated cotton dependencies for elevation, precipitation, and temperature holding two variables constant while varying the three variables given by the x-axis

APPENDIX B

Table B.1. Summary Statistics for Net Returns in Dollars per Acre

	Montana	Colorado	Texas
Mean	127.74	105.35	70.33
Standard Deviation	59.46	34.96	44.37
95 % LCI[1]	47.54	50.88	24.91
95 % UCI[1]	291.48	146.97	116.57
Min	40.82	50.88	24.91
Median	116.57	96.08	59.15
Max	327.01	222.37	268.20
Skewness	1.65	1.36	2.79
Kurtosis	6.64	5.80	13.77

1. LCI is the 95 % lower confidence interval and UCI is the 95% upper confidence interval

Table B.2. Normality Test

	Montana Returns	Colorado Returns	Texas Returns
Chi-Square Statistic	11.72	4.45	11.56
p-Value[1]	0.02	0.35	0.02

1. Null hypothesis is that the data are normally distributed; low p-values indicate rejection of null hypothesis.

Table B.3. Copula Fit Statistics

Copula	SIC[1]	AIC[2]	HQIC[3]
Clayton	0.69	2.15	1.72
Frank	2.05	3.52	3.08
Gaussian	-0.44	3.56	2.65
T	0.25	5.29	4.37
Gumbel	8.59	10.05	9.62

[1] Schwarz information criterion: $SIC = ln[n]k - 2ln[L_{max}]$

[2] Akaike information criterion: $AIC = \left(\frac{2n}{n-k-1}\right)k - 2ln[L_{max}]$

[3] Hannan-Quinn information criterion: $HQIC = 2ln\big[ln[n]\big]k - 2ln[L_{max}]$

Table B.4. Dependency Measures for Pearson, Spearman's Rho, and Copulas

	Texas Returns	Colorado Returns	Montana Returns
Pearson Correlation			
Texas Returns	1	0.351981	0.564219
Colorado Returns		1	0.629165
Montana Returns			1
Rank Correlation/Kendall's Tau			
	Texas Returns	Colorado Returns	Montana Returns
Texas Returns	1	0.104762	0.114286
Colorado Returns		1	0.298413
Montana Returns			1
Rank Correlation based on Clayton Copula			
	Texas Returns	Colorado Returns	Montana Returns
Texas Returns	1	0.1179	0.1649
Colorado Returns		1	0.3718
Montana Returns			1
Rank Correlation based on Frank Copula			
	Texas Returns	Colorado Returns	Montana Returns
Texas Returns	1	0.1812	0.1876
Colorado Returns		1	0.4256
Montana Returns			1
Rank Correlation based on Gaussian Copula			
	Texas Returns	Colorado Returns	Montana Returns
Texas Returns	1	0.1755	0.21776
Colorado Returns		1	0.4489
Montana Returns			1

Table B.5. Efficient Portfolio Statistics

Asymmetric Dependence-Clayton Copula		
Standard Deviation of Returns	CVaR	Expected Return
38.54	-25.16	66.59
33.28	-33.74	72.90
29.09	-40.16	79.21
26.45	-45.37	85.53
25.38	-50.05	91.84
25.53	-54.11	98.16
27.04	-57.25	104.47
29.42	-58.71	110.78
Symmetric Dependence-Pearson Correlation		
Standard Deviation of Returns	CVaR	Expected Return
43.57	21.96	68.32
38.40	5.91	75.04
34.35	-8.14	81.77
31.87	-20.08	88.49
31.33	-28.73	95.22
32.81	-33.98	101.94
35.51	-36.22	108.67

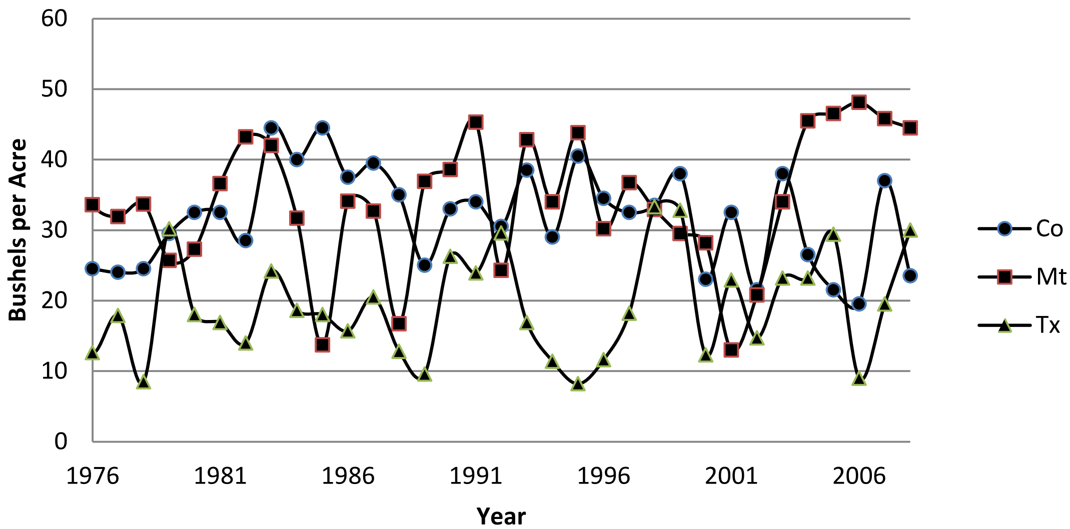

Figure B.1. County level yields per acre

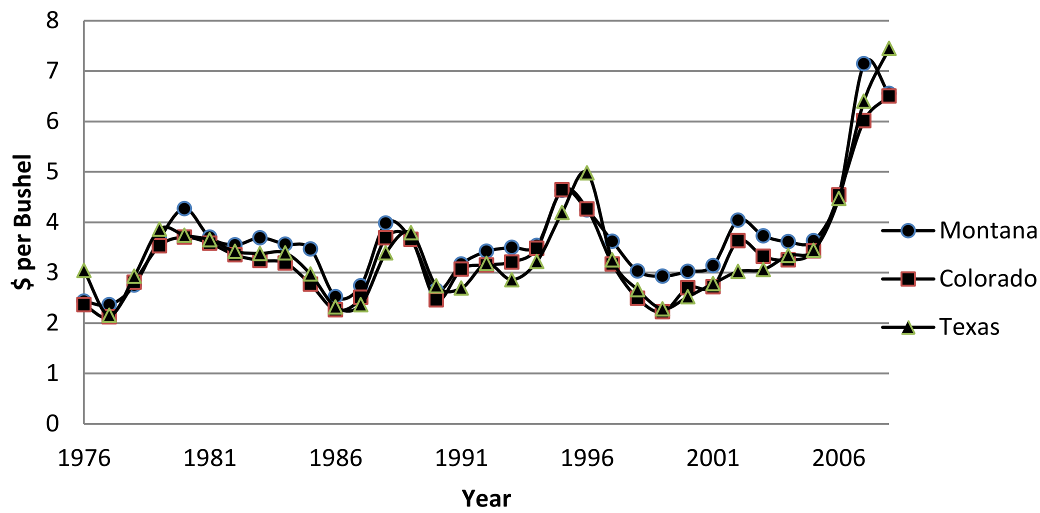

Figure B.2. Price per bushel for wheat

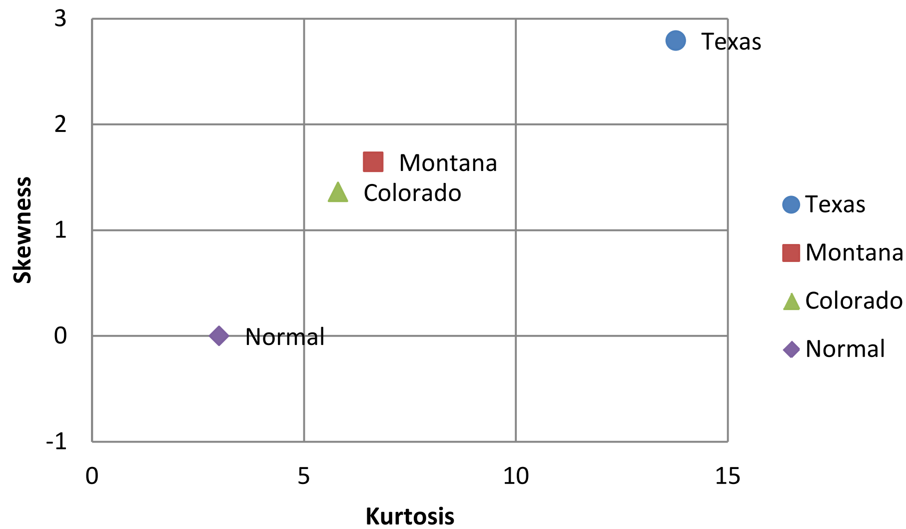

Figure B.3. Skewness and kurtosis for three regions

Figure B.4. Illustration of Clayton, Frank, and Gaussian copulas

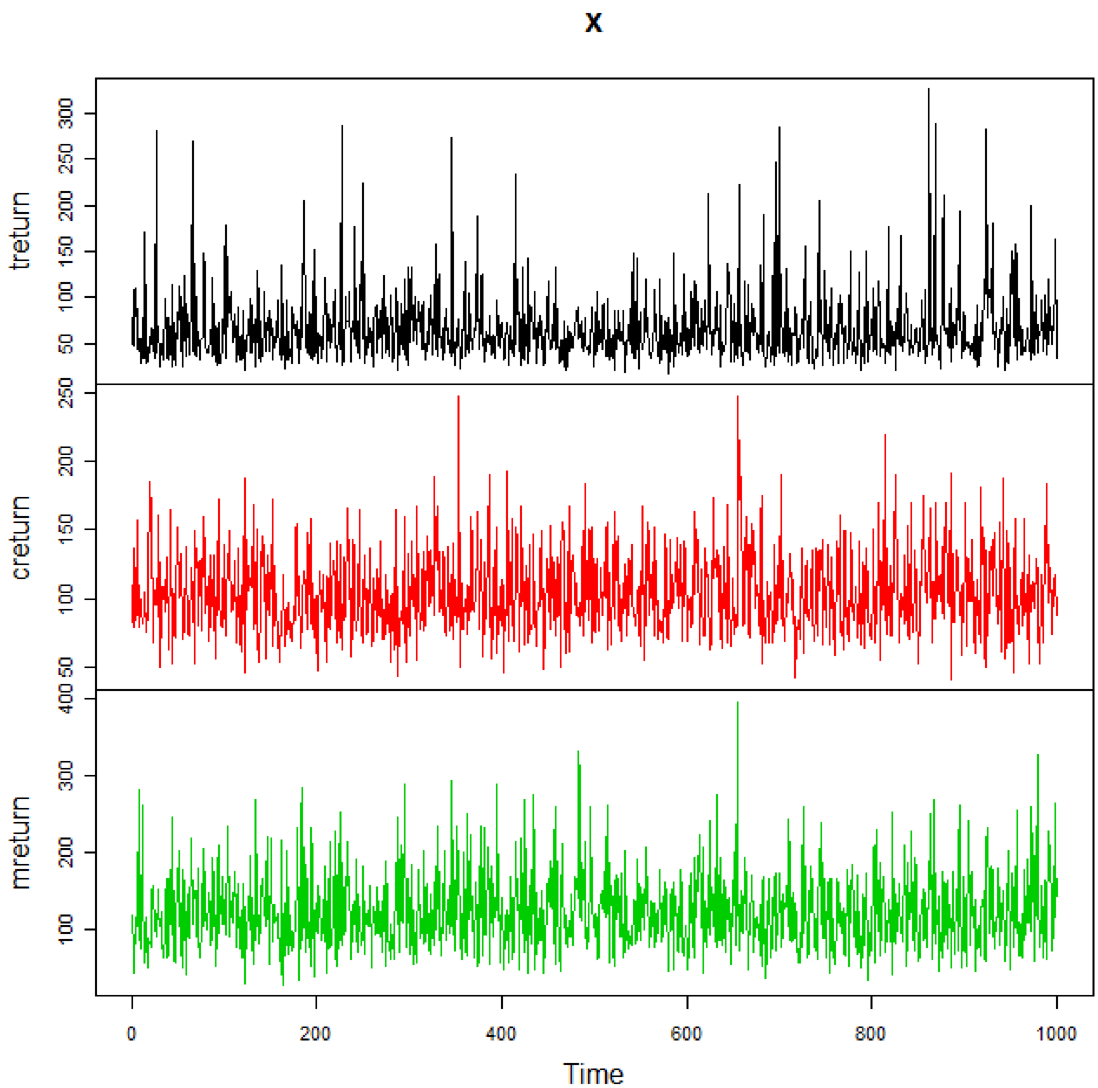

Figure B.5. Clayton copula simulated data

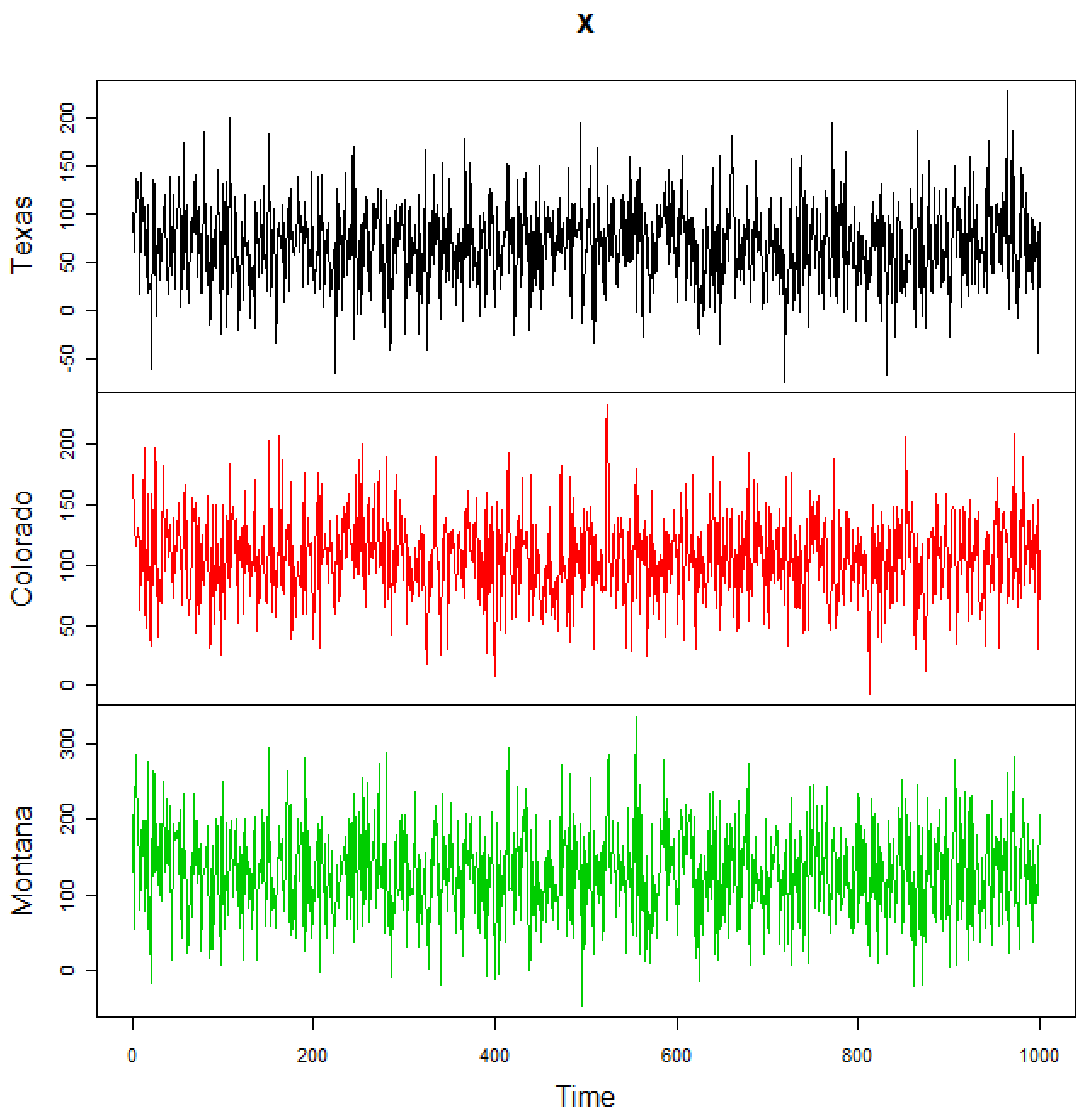

Figure B.6. Multivariate normal simulated data

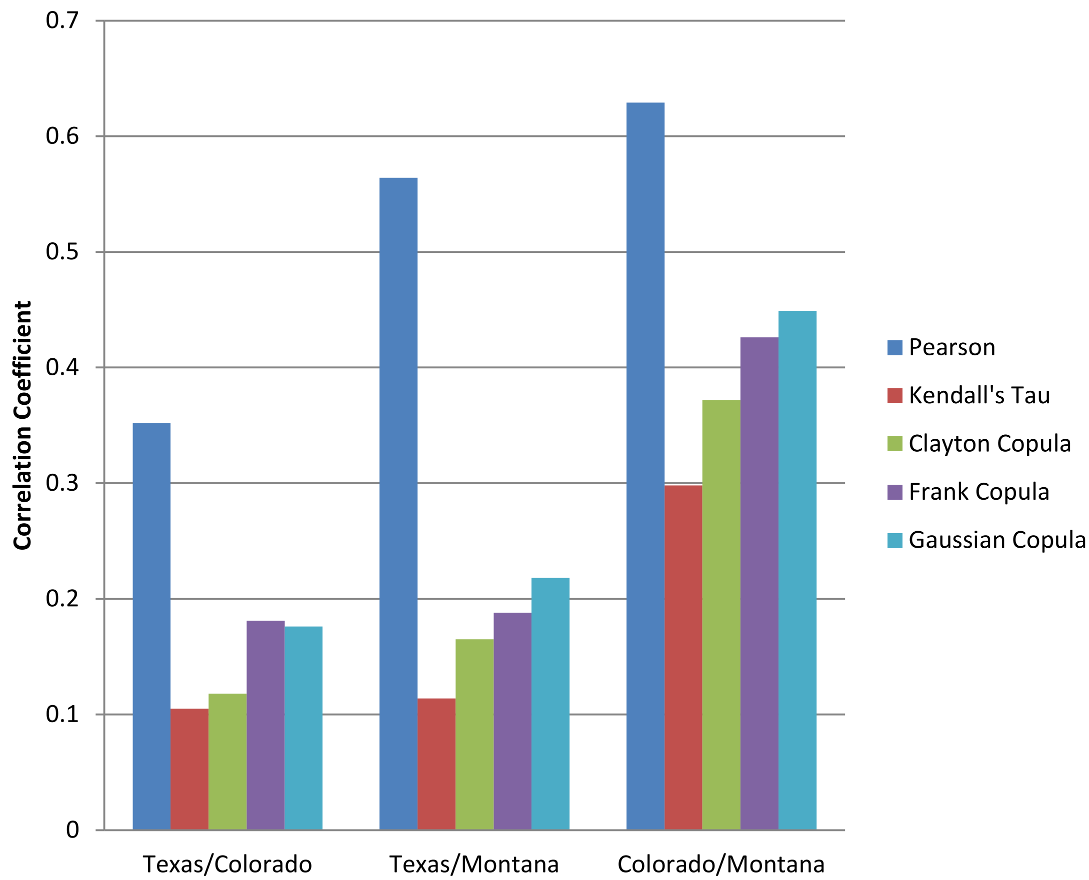

Figure B.7. Dependency measure comparison

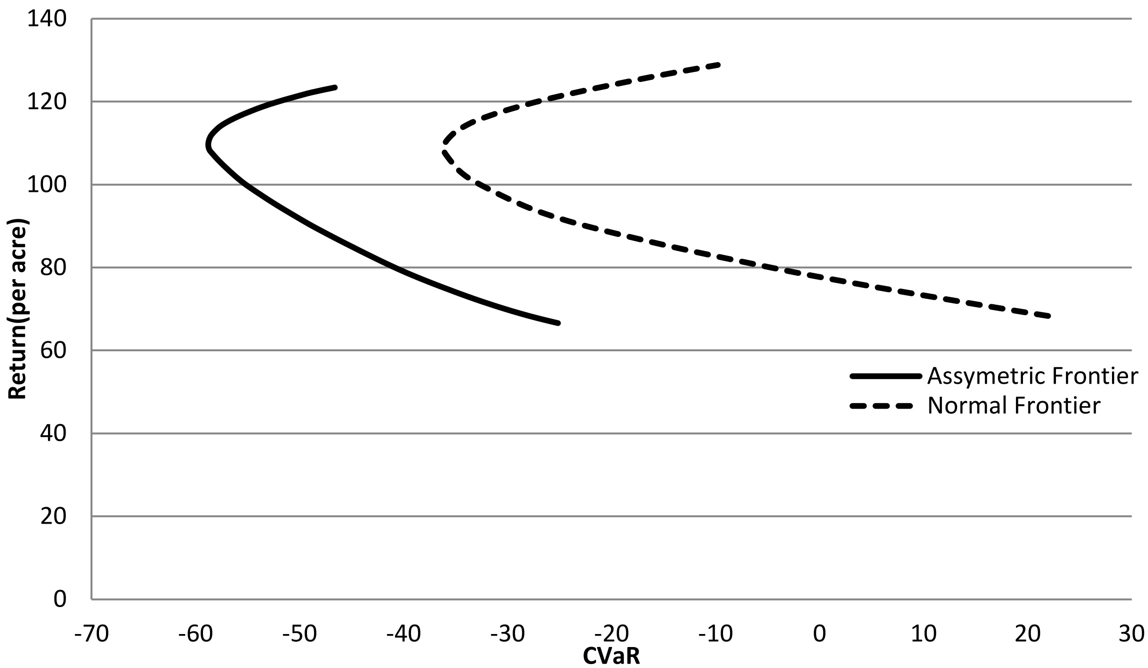

Figure B.8. Asymmetric and normal efficient frontiers

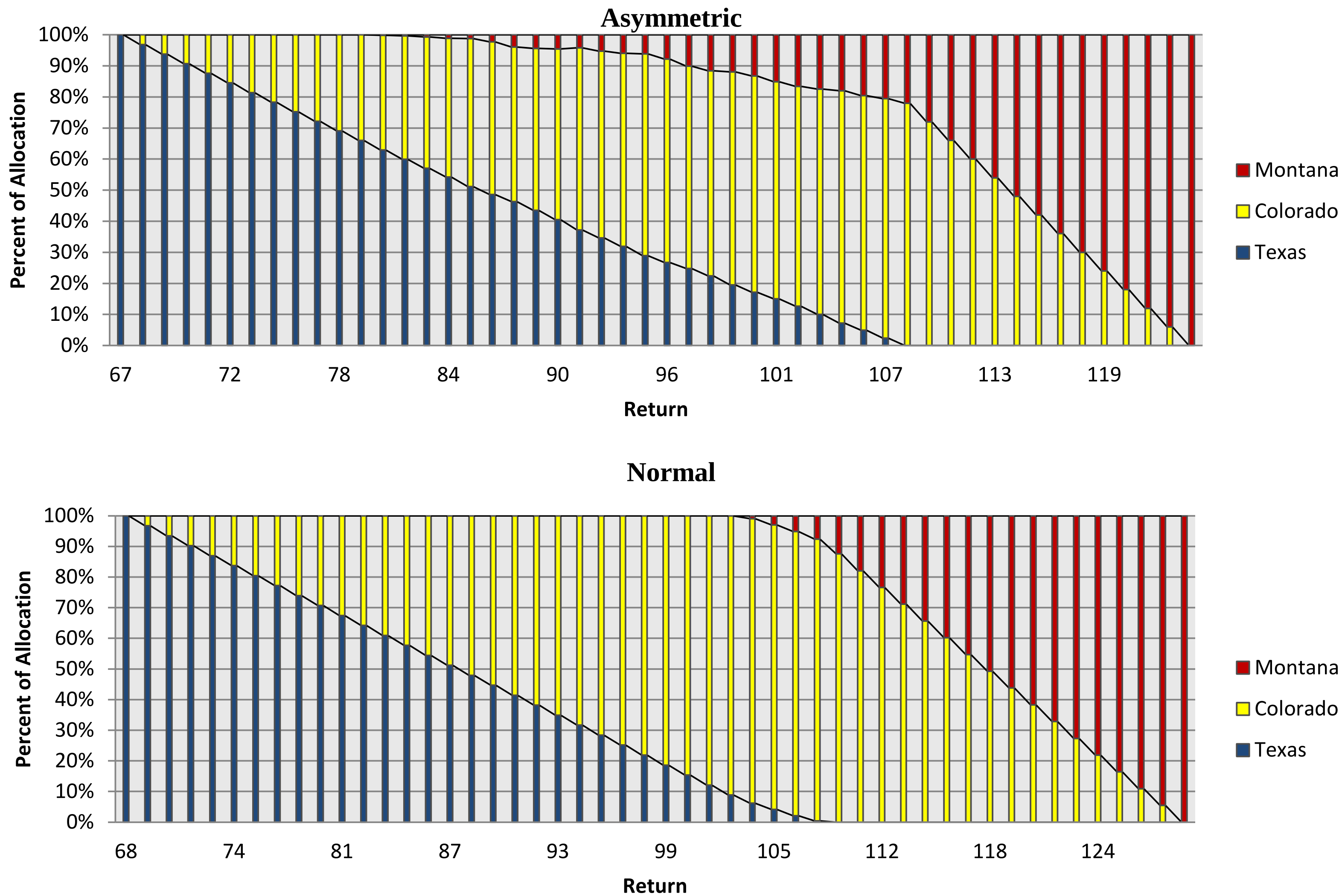

Figure B.9. Acreage allocation between three production regions

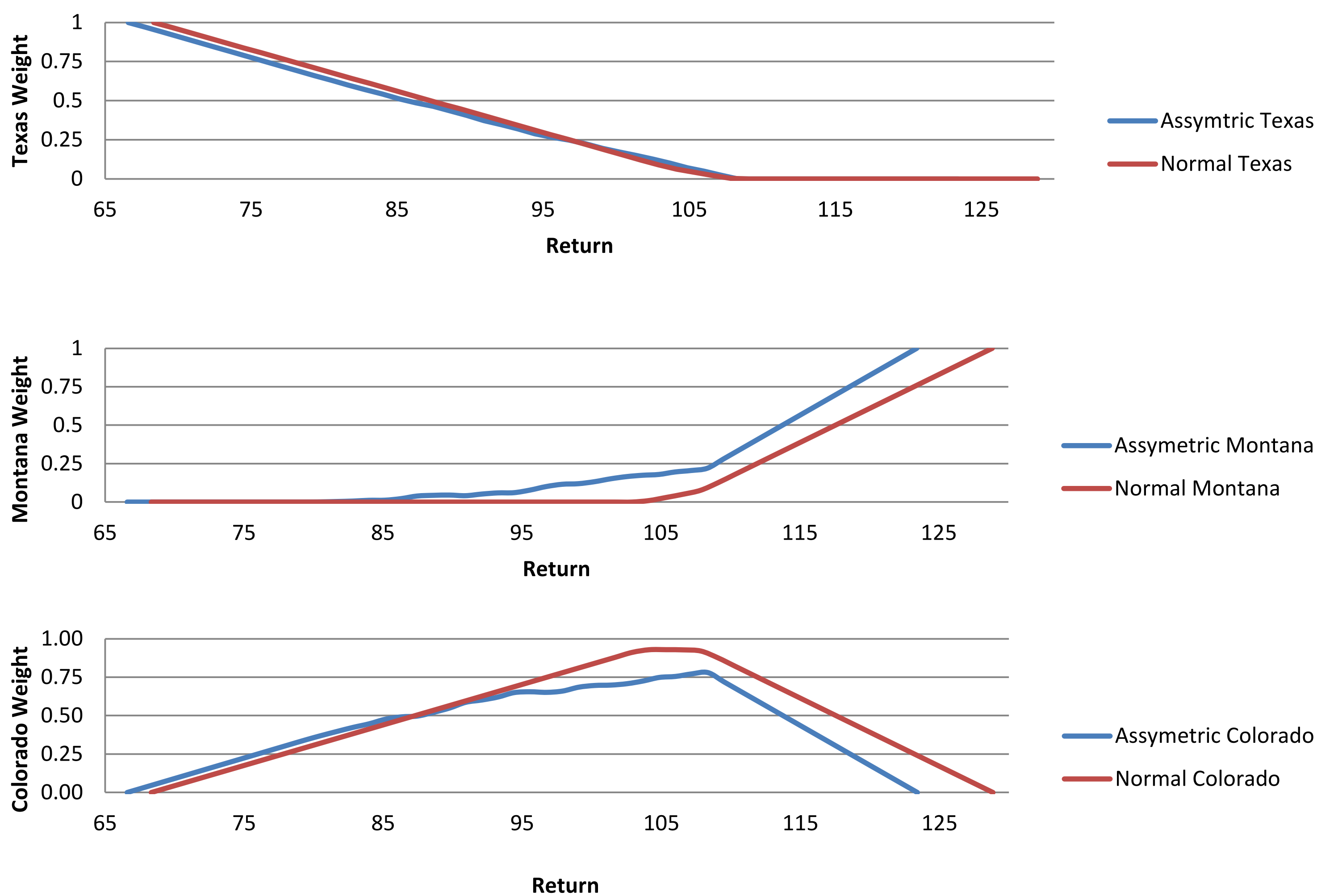

Figure B.4. Acreage allocation comparison for three production regions

APPENDIX C

Table C.1. Historical Data Summary for Period 1974-2008[1]

	Texas Land Price	Colorado Land Price	Montana Land Price	Texas Returns	Colorado Returns	Montana Returns
Mean	394.31	602.62	307.86	70.33	105.35	127.73
St. Deviation	116.62	144.31	184.27	44.37	34.96	59.46
Skewness	2.73	1.83	1.86	2.79	1.36	1.65
Kurtosis	9.22	3.57	3.70	10.77	2.80	3.64

1. Source of data is USDA National Ag Statistics Service and Texas A&M Real Estate Center. Land prices and returns are specified in dollars per acre. Returns are defined as net returns (gross returns less variable cost).

Table C.2. Statistical Summary Based on Annual Change 1974-2008[1]

	Texas Land Price	Colorado Land Price	Montana Land Price	Texas Returns	Colorado Returns	Montana Returns
Mean	16.09	15.69	23.54	5.27	1.68	5.03
St. Deviation	58.51	63.51	72.14	42.47	39.43	44.21
Skewness	0.99	-0.61	-0.73	0.78	1.22	0.15
Kurtosis	2.97	1.22	8.84	0.97	2.75	-0.31

1. Source of data is USDA National Ag Statistics Service and Texas A&M Real Estate Center. Land prices and returns are specified in dollars per acre. Returns are defined as net returns (gross returns less variable cost).

Table C.3. Estimation of Farmland Price Equation[1] Covering the Years 1998-2008

State	Variable	Coefficient Estimate	Standard Error	t-statistic
Texas	Constant	-158.834	53.807	-2.952
	P_{t-1}	1.508**	0.122	12.354
		$R^2 = 0.95, Adjusted\ R^2 = 0.94$		
Colorado	Constant	-17.820	77.300	-0.231
	P_{t-1}	1.080**	0.072	14.932
		$R^2 = 0.97, Adjusted\ R^2 = 0.96$		
Montana	Constant	-12.996	60.572	-0.215
	P_{t-1}	1.165**	0.126	9.222
		$R^2 = 0.91, Adjusted\ R^2 = 0.90$		

1. See equation (8) for land price equation specification.
** Significant at 5% level.

Table C.4. Portmanteau Test for White Noise

	Q-Statistic[1]	ρ-value[2]
Texas Land Price Error Term	2.72	0.44
Colorado Land Price Error Term	2.86	0.42
Montana Land Price Error Term	3.03	0.39

1. $Q = n(n+2)\sum_{j=1}^{m}\frac{1}{n-j}\hat{\rho}(j)$
2. Reject null hypothesis of white noise if ρ-value is less than α.
3. All error terms are unable to reject null hypothesis

Table C.5. Estimation of Gross Return Equation for Three States Covering the Years 1998-2008

State	Variable	Coefficient Estimate	Standard Error	t-statistic
Texas	Constant	0.0653	0.0861	0.76
	R_{t-1}	-0.4421**	0.1504	-2.94
	$\widehat{Var}[\varepsilon_{1t}] = 0.2442, R^2 = 0.213, Adjusted\ R^2 = 0.188$			
Colorado	Constant	2.4899	0.7361	3.38
	R_{t-1}	0.4614**	0.1598	2.89
	$\widehat{Var}[\varepsilon_{1t}] = 0.0806, R^2 = 0.202, Adjusted\ R^2 = 0.178$			
Montana	Constant	1.8796	0.7312	2.57
	R_{t-1}	0.6084**	0.1540	3.95
	$\widehat{Var}[\varepsilon_{1t}] = 0.1337, R^2 = 0.321, Adjusted\ R^2 = 0.301$			

** Significant at 5% level.

Table C.6. Shapiro-Wilk Normality Test[1]

Variable	Mean	St. Deviation	W	z	p value[2]
Colorado Land Residuals	0.00	39.81	0.901	0.756	0.225
Montana Land Residuals	0.00	41.59	0.744	4.396	0.000*
Texas Land Residuals	0.00	53.63	0.960	0.813	0.210
Colorado Returns	0.00	32.98	0.831	3.746	0.000*
Montana Returns	0.00	42.87	1.140	0.273	0.392
Texas Returns	0.00	41.08	0.873	3.147	0.001*

1. $W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$
2. Reject null hypothesis of normally distributed data if p value is less than chosen alpha level.

* Reject null hypothesis of normally distributed data based on 0.05 alpha level.

Table C.7. Skewness/Kurtosis Test for Normality[1]

Variable	Pr(Skewness)	Pr(Kurtosis)	Chi^2	p value
Colorado Land Residuals	0.26	0.95	1.43	0.49
Montana Land Residuals	0.00	0.00	28.45	0.00*
Texas Land Residuals	0.23	0.11	4.13	0.13
Colorado Returns	0.00	0.00	19.12	0.00*
Montana Returns	0.19	0.28	3.13	0.21
Texas Returns	0.00	0.01	15.02	0.00*

1. Based on Chi-square test and reject null hypothesis of normally distributed data if p value is less than chosen alpha level.

* Reject null hypothesis of normally distributed data based on 0.05 alpha level.

Table C.8. Correlation Matrix for Change in Land Prices and Net Revenues[1] for 1974-2008

	Texas Land	Colorado Land	Montana Land	Texas Returns	Colorado Returns	Montana Returns
Texas Land	1.00	0.92	0.37	0.24	0.05	-0.12
Colorado Land	0.92	1.00	0.38	0.15	0.01	-0.09
Montana Land	0.37	0.38	1.00	0.12	0.26	0.23
Texas Returns	0.24	0.15	0.12	1.00	0.05	-0.24
Colorado Returns	0.05	0.01	0.26	0.05	1.00	0.32
Montana Returns	-0.12	-0.09	0.23	-0.24	0.32	1.00

1. Net revenues are defined as the gross revenue less variable cost.

Table C.9. Copula Akaike Information Criteria Fit Statistics[1]

Copula	Texas Returns/ Montana Returns	Texas Returns/Colorado Returns	Texas Returns/Texas Land Price	Colorado Returns/Montana Returns	Colorado Returns/Texas Land Price	Colorado Returns/Colorado Land Price	Colorado Returns/Montan a Land Price	Montana Returns/Texas Land Price
Gaussian	2.84	0.98	-1.64	2.12	-2.42	-2.47	-2.49	-2.10
T	1.43	-0.62	-3.12	2.32	-5.18	-5.72	-5.66	-4.73
Clayton	0.90	-1.84	-4.01	0.99	-5.60	-5.74	-5.68	-4.83
Gumbel	-0.08	-2.10	-4.72	-0.20	-5.62	-5.75	-5.68	-5.01
Frank	-0.16	-2.74	-4.94	-1.14	-5.67	-5.68	-5.71	-5.30

Copula	Texas Returns/ Colorado Land Price	Texas Returns/Montana Land Price	Texas Returns/Texas Land Price	Montana Returns/Colorado Land Price	Montana Returns/Montana Land Price	Texas Land Price/Colorado Land Price	Texas Land Price/Montana Land Price	Colorado Land Price/Montana Land Price
Gaussian	-1.58	-1.91	-1.64	-1.75	-2.36	4.33	2.95	-0.56
T	-4.71	-4.61	-3.12	-4.99	-5.59	1.73	0.67	-3.14
Clayton	-4.28	-4.86	-4.01	-5.16	-5.59	3.06	-1.62	-2.47
Gumbel	-4.46	-4.93	-4.72	-5.00	-5.66	2.63	1.14	-2.88
Frank	-4.92	-5.18	-4.94	-5.01	-5.63	0.81	-0.47	-4.09

1. The values in the table represent the AIC fit statistics. A better fit is represented by lower values.

Table C.10. Historical Data Compared to Scenario Generated Data for First Two Stages in Five Stage Model

	Node 1		Node 2		Node 3		Node 4		Node 5		Node 6	
	Historical	Scenario	Historical	Scenario	Historical	Scenario	Historical	Scenario	Historical	Scenario	Historical	Scenario
	Mean											
Texas Land Price Error Term	0.00	0.01	0.00	0.00	0.00	-0.021	0.00	0.01	0.00	0.00	0.00	0.00
Colorado Land Price Error Term	0.00	0.00	0.00	0.00	0.00	-0.030	0.00	0.01	0.00	0.00	0.00	0.00
Montana Land Price Error Term	0.00	0.02	0.00	0.00	0.00	0.016	0.00	-0.01	0.00	0.00	0.00	0.00
Texas Gross Returns	102.24	102.24	97.49	97.49	101.24	101.26	111.61	111.61	98.63	98.63	98.63	98.63
Colorado Gross Returns	106.26	106.28	103.59	103.59	115.21	115.21	107.37	107.37	103.70	103.70	103.70	103.70
Montana Gross Returns	159.32	159.32	149.73	149.73	172.55	172.54	164.18	164.18	153.60	153.60	153.60	153.60
	Standard Deviation											
Texas Land Price Error Term	37.42	37.42	37.11	37.11	35.31	35.29	37.59	37.59	35.33	35.33	35.71	35.71
Colorado Land Price Error Term	37.77	37.78	48.67	48.67	39.30	39.30	41.81	41.81	44.62	44.61	36.98	36.98
Montana Land Price Error Term	61.76	61.76	72.76	72.76	69.14	69.17	69.03	69.03	69.14	69.13	90.09	90.08
Texas Gross Returns	63.96	63.96	65.87	65.87	61.06	61.07	62.13	62.13	67.81	67.81	62.04	62.04
Colorado Gross Returns	46.22	46.22	47.06	47.06	52.37	52.37	44.53	44.53	44.21	44.20	44.80	44.80
Montana Gross Returns	89.87	89.87	95.33	95.33	95.35	95.35	85.91	85.91	87.06	87.05	87.57	87.57
	Skewness											
Texas Land Price Error Term	-0.793	-0.714	-0.806	-0.808	-0.659	-0.649	-0.180	-0.187	-0.652	-0.640	-0.924	-0.822
Colorado Land Price Error Term	0.752	0.703	0.390	0.443	0.741	0.752	0.631	0.648	0.477	0.556	0.686	0.659
Montana Land Price Error Term	2.522	2.236	1.234	1.277	1.225	1.264	1.234	1.264	1.225	1.242	0.428	0.619
Texas Gross Returns	1.863	1.976	1.962	2.023	1.969	2.102	1.996	2.084	1.283	1.227	2.000	2.136
Colorado Gross Returns	1.803	1.731	1.939	1.906	1.207	1.114	1.947	1.914	1.755	1.757	1.942	1.879
Montana Gross Returns	0.704	0.596	0.813	0.694	0.415	0.519	0.799	0.711	0.521	0.514	0.869	0.626
	Kurtosis											
Texas Land Price Error Term	1.344	1.274	1.856	1.784	1.853	1.769	0.996	0.954	1.847	1.777	1.532	1.547
Colorado Land Price Error Term	-0.174	-0.121	-0.283	-0.257	0.284	0.306	-0.010	0.011	-0.219	-0.210	0.399	0.461
Montana Land Price Error Term	6.967	4.999	2.176	2.163	2.145	2.123	2.175	2.128	2.147	2.057	0.423	0.335
Texas Gross Returns	4.076	4.080	4.454	4.368	4.695	4.584	4.617	4.536	1.146	1.147	4.637	4.643
Colorado Gross Returns	3.128	3.096	3.692	3.702	0.133	0.160	3.756	3.767	3.223	3.224	3.706	3.685
Montana Gross Returns	-0.553	-0.529	-0.412	-0.392	-1.307	-1.341	-0.182	-0.161	-0.658	-0.667	-0.246	-0.208

Table C.11. Optimal Acreage with Land Only in One Region

	Mean	Min	Max	SD
		Texas		
Stage 0	3000	3000	3000	0
Stage 1	4500	4500	4500	0
Stage 2	4775	4627	5092	170
Stage 3	5012	4674	5558	194
Stage 4	4967	4513	5792	168
Stage 5	4864	4500	6099	297
		Colorado		
Stage 0	3000	3000	3000	0
Stage 1	4500	4500	4500	0
Stage 2	4284	4193	4500	120
Stage 3	4582	4429	4950	132
Stage 4	4617	4479	4986	100
Stage 5	4634	4500	5337	155
		Montana		
Stage 0	3000	3000	3000	0
Stage 1	4500	4500	4500	0
Stage 2	4082	3913	4283	129
Stage 3	3869	2956	4797	572
Stage 4	3203	2051	4770	708
Stage 5	2757	2000	6045	960

Table C.12. Optimal Acreage Allocations Given Geographical Diversification

	Colorado				Texas				Montana			
	Mean	Min	Max	SD	Mean	Min	Max	SD	Mean	Min	Max	SD
					Texas Base Region							
Stage 0	250	250	250	0	3000	3000	3000	0	250	250	250	0
Stage 1	1332	1332	1332	0	4408	4408	4408	0	1750	1750	1750	0
Stage 2	628	332	1321	404	4064	3408	4772	568	952	750	1713	413
Stage 3	951	250	1875	482	3783	2450	4814	488	1015	250	1228	413
Stage 4	818	250	3104	707	4310	2000	5000	782	371	250	1228	227
Stage 5	1020	250	3589	996	3896	2000	5000	918	258	250	1145	70
					Colorado Base Region							
Stage 0	3000	3000	3000	0	250	250	250	0	250	250	250	0
Stage 1	3000	3000	3000	0	1750	1750	1750	0	1750	1750	1750	0
Stage 2	2406	2000	3000	385	1888	750	2565	684	1118	750	1706	499
Stage 3	2594	2000	3454	446	1877	297	2809	601	989	250	1300	414
Stage 4	2664	2000	4954	901	2262	250	3306	968	418	250	1300	281
Stage 5	2978	2000	5000	1066	1905	250	3736	1007	263	250	1379	95
					Montana Base Region							
Stage 0	250	250	250	0	250	250	250	0	3000	3000	3000	0
Stage 1	1750	1750	1750	0	1750	1750	1750	0	4500	4500	4500	0
Stage 2	1152	750	1866	421	2532	750	3250	956	3750	3500	4000	396
Stage 3	1712	250	2250	586	2923	2000	4040	762	3176	2500	3000	481
Stage 4	1193	250	3540	446	3053	1250	4748	545	2578	2000	3000	358
Stage 5	944	250	5000	960	2680	250	5000	851	2012	2000	3344	105

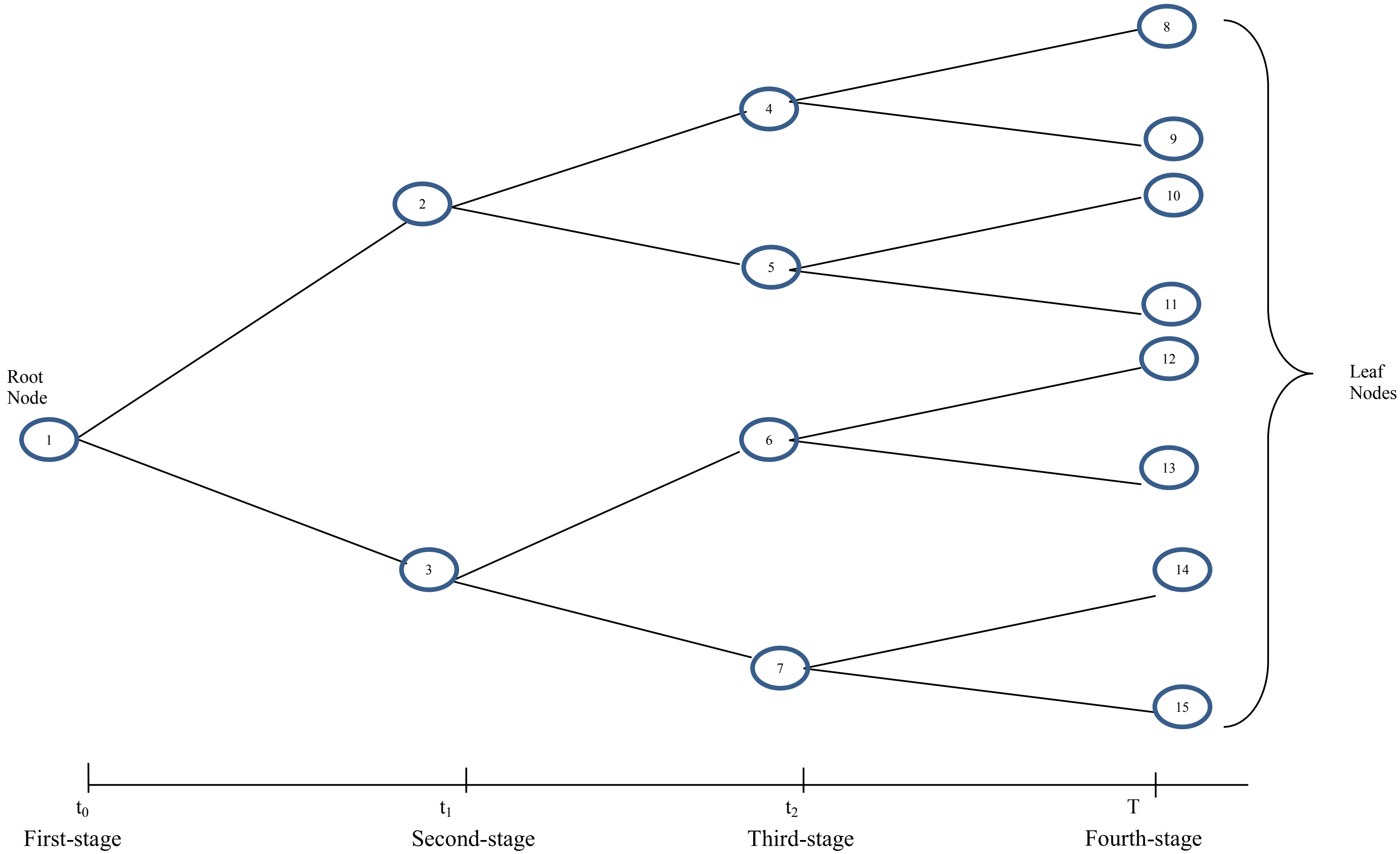

Figure C.1. Scenario tree[4]

[4] An important concept to understand with a scenario tree is the concept of the father node. The node $a(i)$ is referred to as the *father* of node i. For example, the father of node 2 is node 1, or $a(2) = 1$.

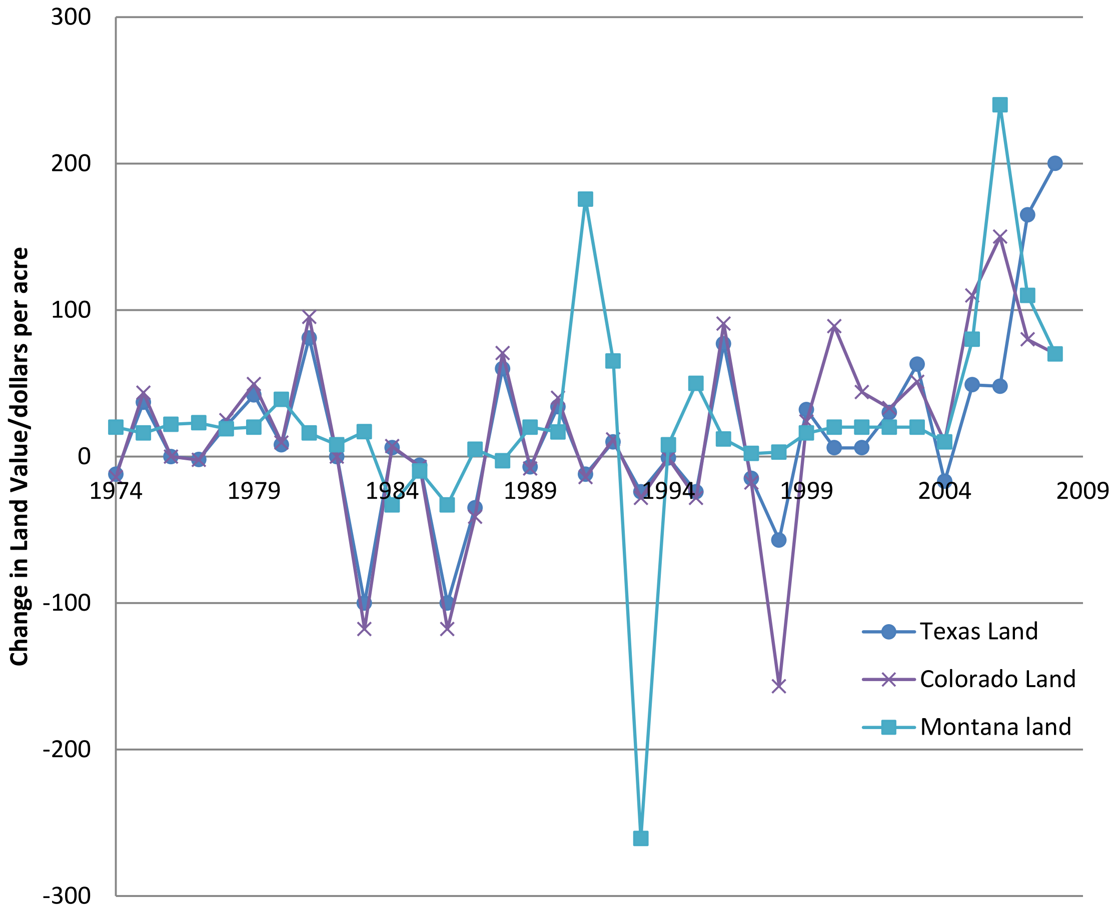

Figure C.2. Change in land value 1974-2008[5]

[5] The sources of the data are USDA and Texas A&M Real Estate Center.

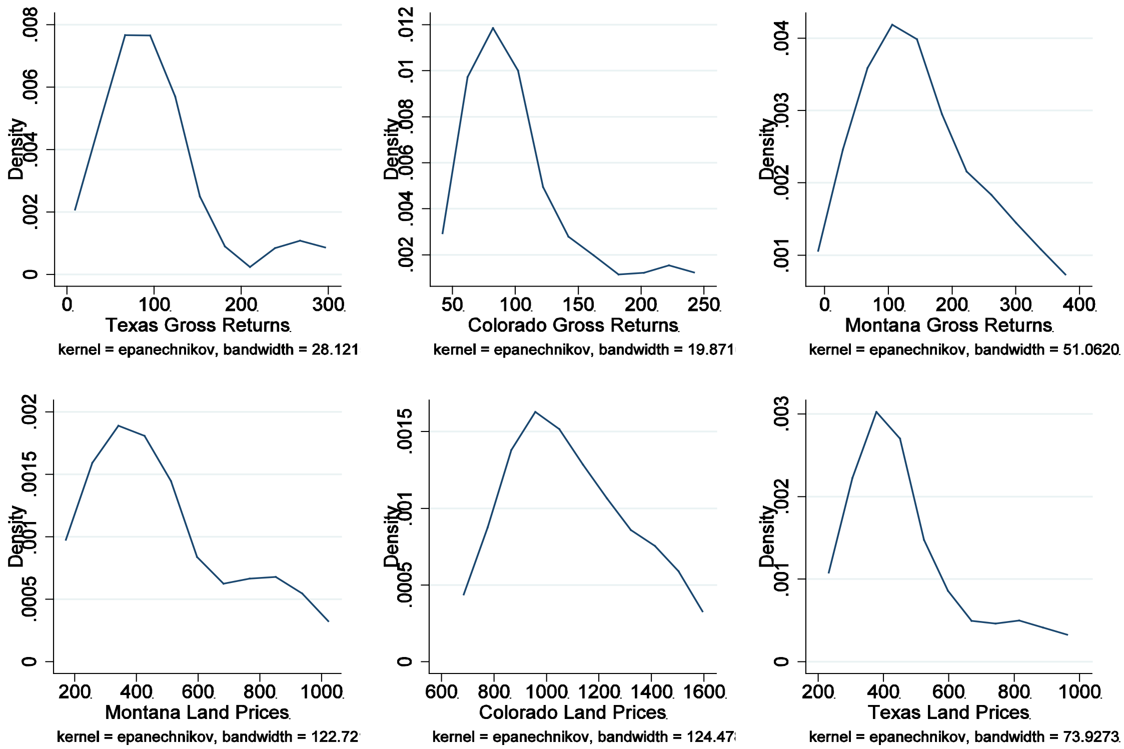

Figure C.3. Kernel estimated pdf's

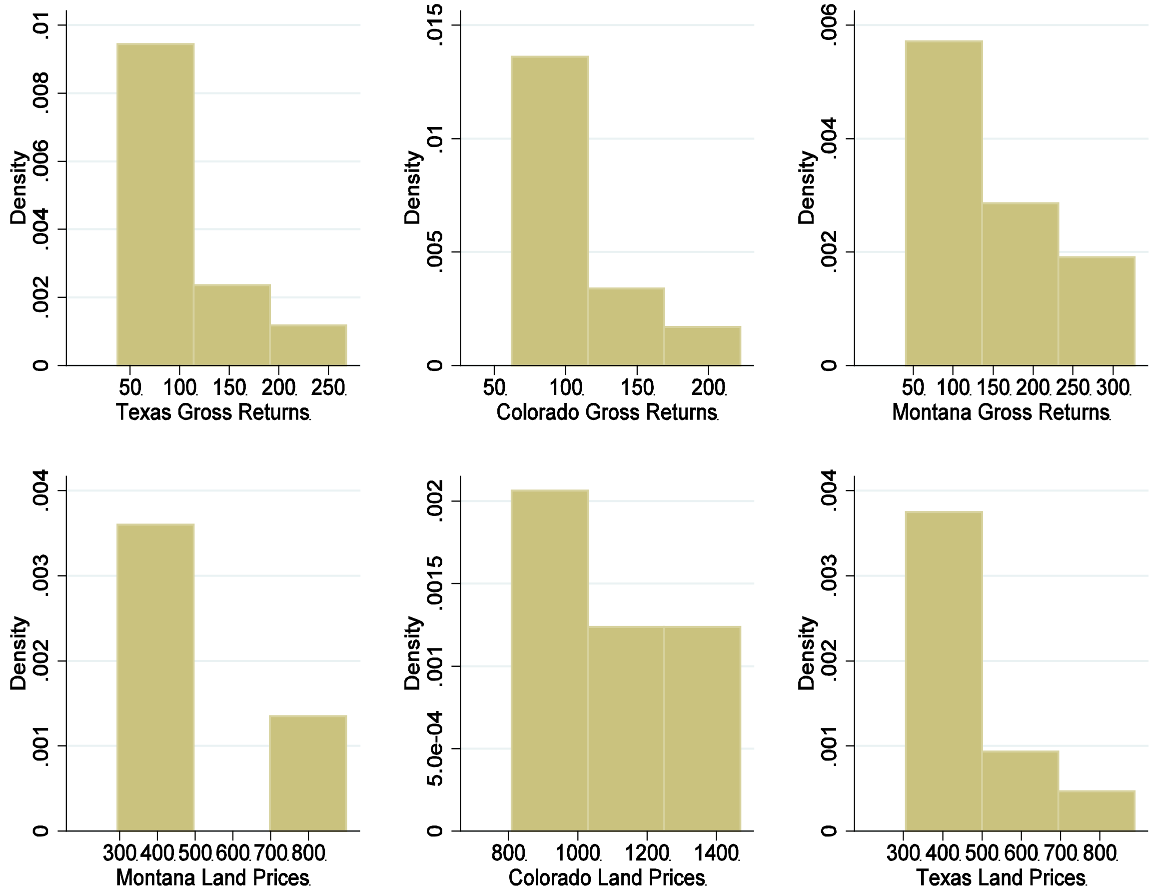

Figure C.4. Histogram for gross returns and land prices[6]

[6] Data covers the time period 1998-2008 and is based off the historical data.

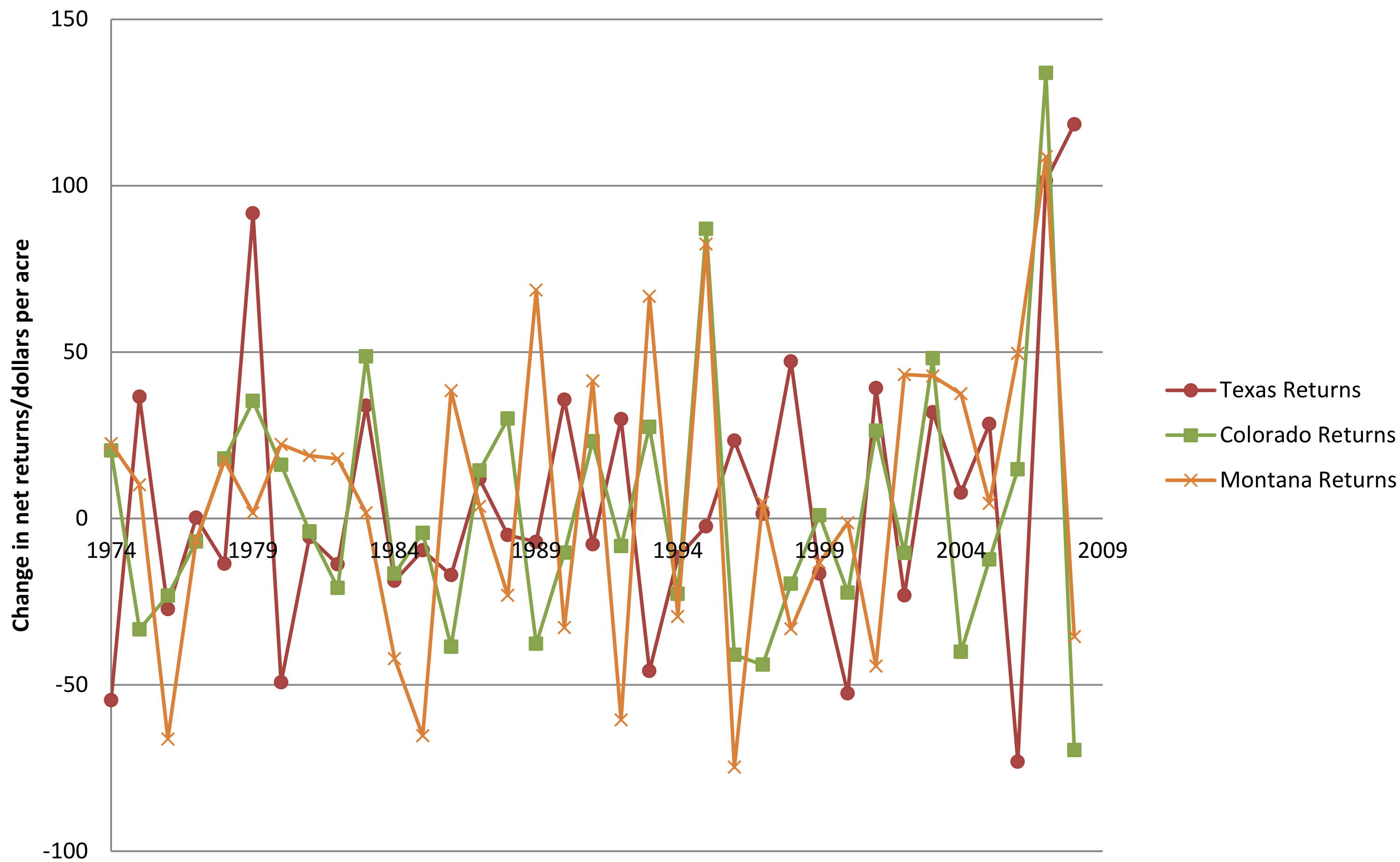

Figure C.5. Change in wheat net returns for three growing regions 1974-2008

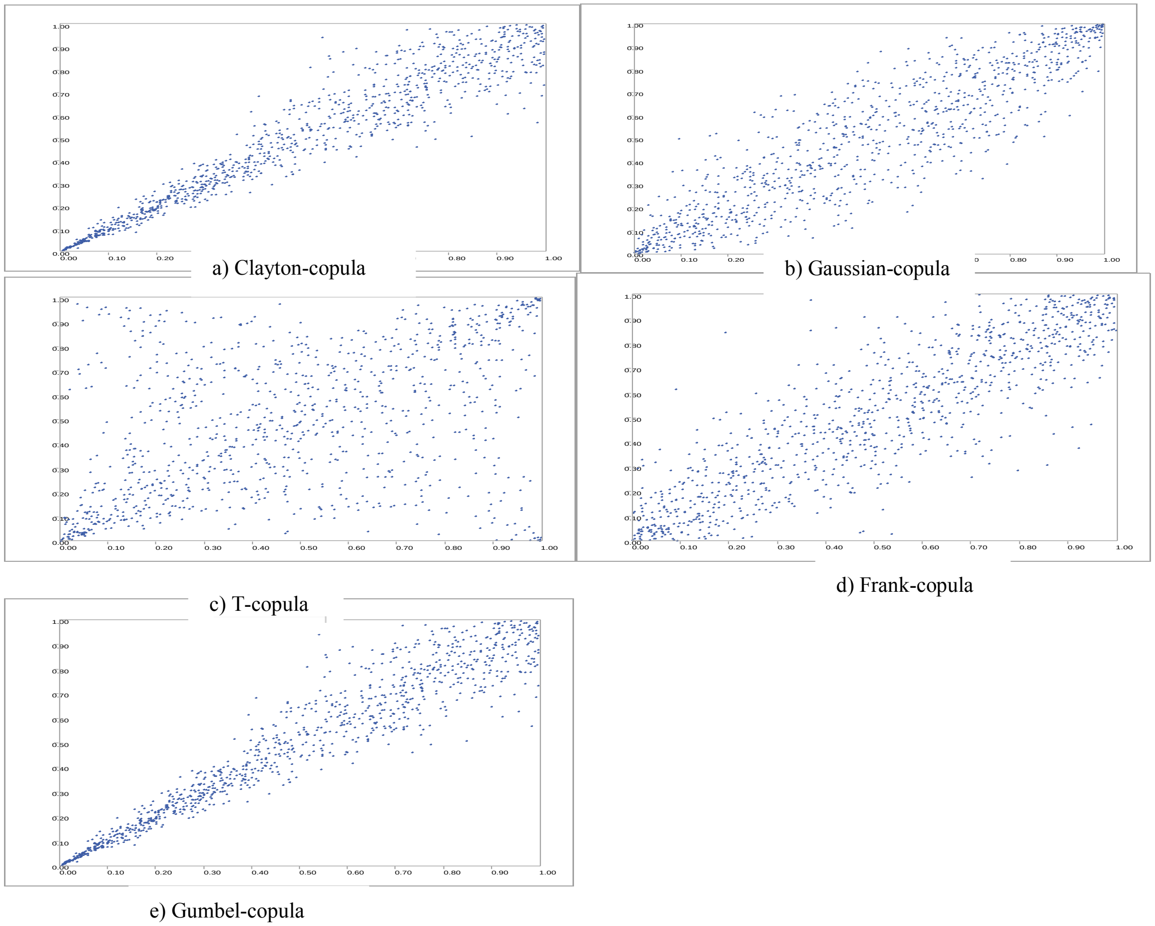

Figure C.6. Illustration of the shape differences between five copulas[7]

[7] AIC fit criteria showed that the Gaussian copula provided the best fit.

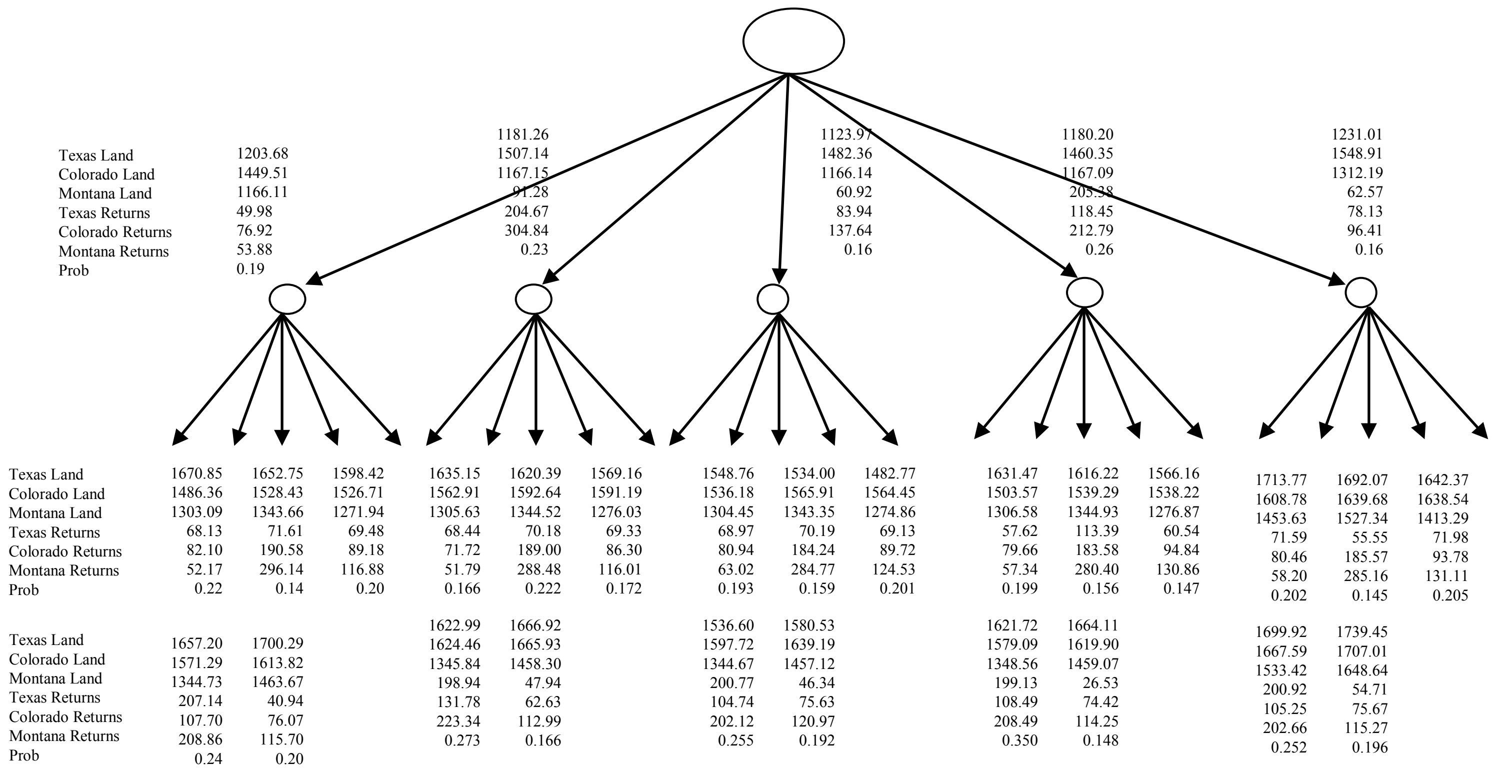

Figure C.7. The first two stages of the generated five-stage scenario tree[8]

[8] The last two observations for each node are shown as overflows in the bottom row.

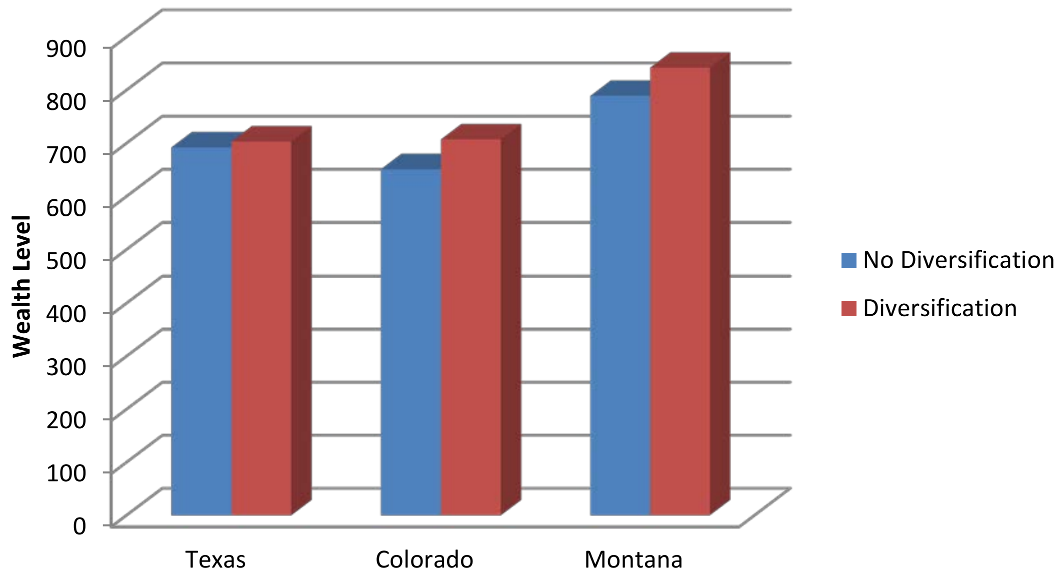

Figure C.8. Certainty equivalents under different diversification assumptions

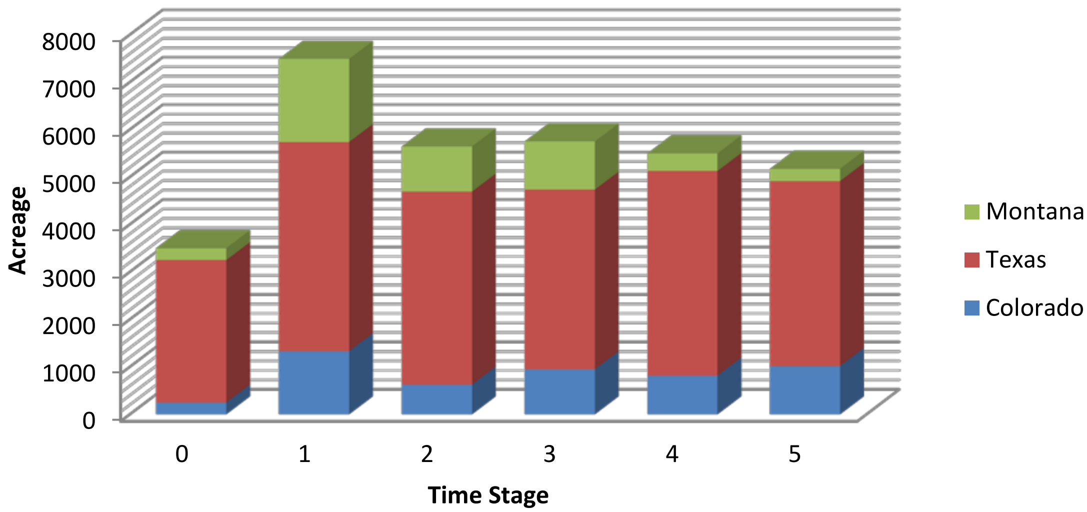

Figure C.9. Acreage allocations with Texas as base region

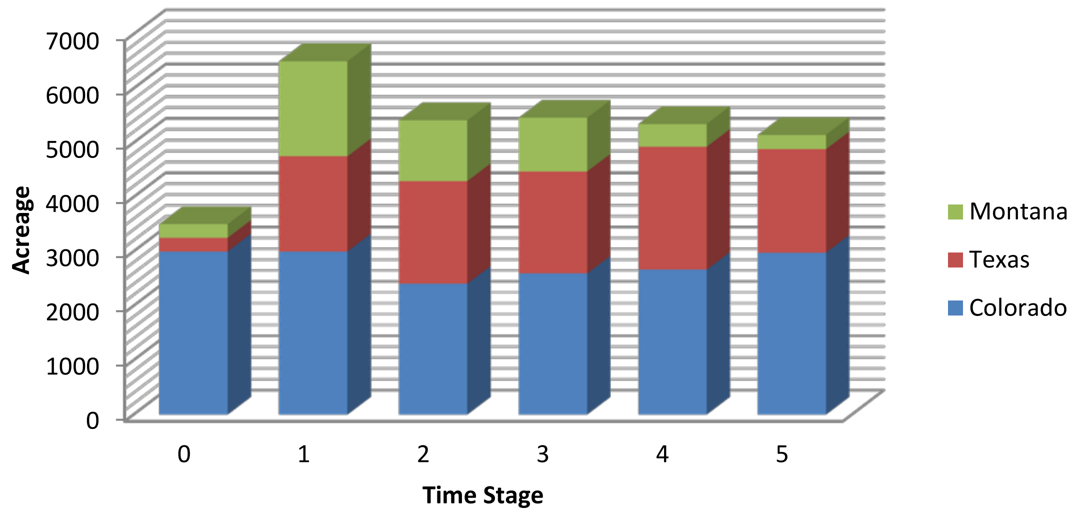

Figure C.10. Acreage allocations with Colorado as base region

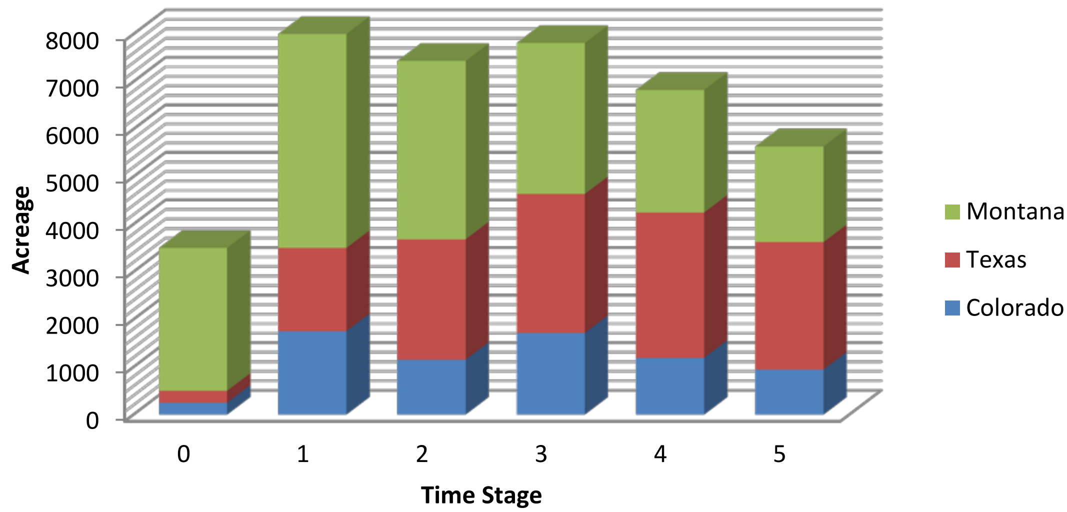

Figure C.11. Acreage allocations with Montana as base region

APPENDIX D

DISCRETE STOCHASTIC PROGRAMMING FORMULATION

The expected utility of terminal wealth is maximized subject to land, debt, and financial constraints. A description of each constraint both mathematically and verbally is provided in this section. An understanding of the subscripts is important in describing the mathematical representations of the constraints. The subscript r denotes the region where production occurs. Texas, Colorado, and Montana are the three regions denoted by r. The subscript t denotes the year in which the decisions are made with T being the terminal Stage. The subscript i denotes the node at stage t which implies that $i = 1,\ldots, I_t$ and I_t is the number of nodes at time t. The subscript j denotes the node at stage t-1 which implies that $j = 1,\ldots,I_{t-1}$ and I_{t-1} is the number of nodes at time t-1.

Land Constraints

One of the most logical constraints that exist in farming is the land constraint. A farmer cannot raise crops on land that he does not own or rent. The first constraint is meant to restrict production to previously owned ground or newly rented or purchased ground

$$\sum_{i=1}^{r}(-IL_r - PL_{11r} + SL_{11r} + Rw_{11r}) \leq 0 \tag{1a}$$

$$\sum_{i=1}^{r}\left(-L_{t-1jr} - PL_{tir} + SL_{tir} + Rw_{tir}\right) \leq 0 \tag{1b}$$

Equation 1a is the initial land constraint at the root node. It specifies that the initial land (IL) in region r less purchased land (PL) in region r plus land sold (SL) in region r, land planted to wheat (Rw) in region r, is less than or equal to zero. Equation 1b extends the first stage equation to the subsequent stages (*t,i*), by constraining that the land owned at the end of the previous Stage (*L*), less purchased land (*PL*) plus sold land (*SL*), plus land used for production is less than zero.

$$-IL - PL_{11r} + SL_{11r} + L_{11r} = 0 \quad (2a)$$

$$-L_{t-1jr} - PL_{tir} + SL_{tir} + L_{tir} = 0 \quad (2b)$$

Equations 2a and 2b are used to transfer land from one Stage to the next. These are considered to be the land accounting constraints.

Machinery Constraints

One of the key constraints when dealing with land allocation is the issue of machinery capacity. It is assumed that there is a given amount of machinery that can be used for a given amount of acreage. Equation 3a accounts for the initial land (*IL*), the acreage that has machinery (*PM*), and the acreage that requires machinery (*AM*). Equation 3b is essentially the same but accounts for machinery depreciation (*dm*) for the machinery used in the last Stage.

$$-IL - PM_{11} + AM_{11} = 0 \quad (3a)$$

$$-dmAM_{t-1j} - PM_{ti} + AM_{ti} = 0 \quad (3b)$$

Equations 4a and 4b are machinery accounting constraints. They are used to transfer machinery value (M_{ti}) from one Stage to the next. Once again depreciation of machinery is accounted for in subsequent Stages.

$$-IM - maPM_{11} + M_{11} = 0 \quad (4a)$$

$$-dmM_{t-1j} - maPM_{ti} + M_{ti} = 0 \quad (4b)$$

Equations 5a and 5b are used to control the amount of machinery purchased. They ensure that the amount of machinery owned and newly purchased is sufficient to service the acreage in production($\sum_{i=1}^{r} W_{11r}$).

$$-IL + \sum_{i=1}^{r} W_{11r} - PM_{11} \leq 0 \tag{5a}$$

$$-dmAM_{t-1j} + \sum_{i=1}^{r} W_{tir} - PM_{ti} \leq 0 \tag{5b}$$

Financial Constraints

These constraints are used to control the debt available to the farmer. The debt level is both a function of the weighting of credit capacity of the farmer (e_{ti}) and the owners equity (OE_{ti}). After the first Stage, this constraint is also a function of credit capacity (d_{ti}) and previous year debt (D_{t-1j}) .

$$-e_{ti}OE_{11} + D_{11} \leq 0 \tag{6a}$$

$$-d_{ti}D_{t-1j} - e_{ti}OE_{ti} + D_{ti} \leq 0 \tag{6b}$$

Equation 7 is used as the accounting constraint for the problem. It represents the fundamental accounting constraint, assets equal liabilities plus owner's equity.

$$pl_{ti}L_{ti} + M_{ti} + \sum_{i=1}^{r} ww_{tir}W_{tir} - OE_{ti} - D_{ti} = 0 \tag{7}$$

Equations 8a and 8b are used to calculate the owner's equity. Recall that ending owner's equity is being used as the measure of terminal wealth. 8a represents the first Stage calculation of owner's equity where *IOE* is the initial owners' equity, *tc* represents the transaction cost of selling land, *SL* is the land sold in each region *r*, and *OE* is the actual owner's equity at time Stage 1. 8b is the constraint used to calculate owner's equity in the following Stages[9]. The profit from growing wheat in each region r is represented by $\sum_{i=1}^{r} prw_{tir} W_{t-1jr}$. The depreciation of machinery is represented by$admM_{t-1j}$, where adm is the after tax depreciation rate and M is the value of machinery. The interest charges from debt are accounted for with $rint_{ti}D_{t-1j}$, where $rint$ is the interest rate and D is the debt level. The capital gains from changes in land values are captured by the term $cg_{tir}L_{t-1jr}$, where cg represents the capital gain on the land.

$$IOE - tc_{11}SL_{11r} - OE_{11} = 0 \quad (8a)$$

$$\sum_{i=1}^{r} prw_{tir}W_{t-1jr} - admM_{t-1j} - rint_{ti}D_{t-1j} + OE_{t-1j} +$$

$$cg_{tir}L_{t-1jr} - tc_{ti}SL_{ti} - OE_{ti} = c \quad (8b)$$

[9] Family withdrawals are represented by the variable c, which are assumed to be constant over the planning Stages.

VITA

Ryan A. Larsen

Contact Information

Assistant Professor
NDSU Dept 7610
540 Barry Hall
North Dakota State University
Fargo, ND 58108-6050
(701) 231-5747
(701) 231-7400 fax
ryan.larsen@ndsu.edu

Education

Ph.D., Agricultural Economics, August 2011,. , Texas A&M University, College Station, Texas

M.S., Agricultural Economics, December 2006, University of Idaho, Moscow, Idaho

B.A., Finance, August 2003, Idaho State University, Pocatello, Idaho

Academic Experience

Assistant Professor — April 2010-Present
Department of Agribusiness and Applied Economics
North Dakota State University
Fargo, North Dakota

Graduate Assistant — Summer 2006-April 2010
Department of Agricultural Economics
Texas A&M University
College Station, Texas

Research Associate — Spring 2004-Summer 2005
Department of Agricultural Economics
University of Idaho
Moscow, Idaho