Abstract

DiRienzo, Cassandra Elizabeth. An Exploration of the Spatial Dependence Structure of Crop Yields and the Implications for Crop Insurance. (Under the direction of Paul Fackler and Barry Goodwin.)

Systemic risk, in regard to agricultural production, refers to the spatial dependence of crop yields stemming from correlated weather, soil patterns, and other geographically related factors. Systemic risk has been named as a contributing factor to the Federal Crop Insurance Corporation's (FCIC) poor actuarial performance. To date, little research has explored the spatial dependence structure of crop yields. This paper explores three aspects of the spatial dependence structure of yields; the cross-crop spatial correlation structure, the rate of crop yield spatial correlation decay, and the characteristics of the bivariate distributions which define the spatial relationship of crop yields.

This paper is divided into six sections. Section One provides a history of the FCIC and a literature review concerning the poor actuarial performance of the FCIC. Section Two discusses the data used in the analysis sections of this paper. Section Three uses two non-parametric tests to explore cross-crop spatial correlation. Section Four develops a model to describe the rate of crop yield spatial correlation decay. Section Five uses the copula methodology to find the bivariate copula distribution that best describes the spatial relationship of crop yields. Section Six summarizes and concludes the research performed in this paper.

An Exploration of the Spatial Dependence Structure of Crop Yields and the Implications for Crop Insurance

by

Cassandra Elizabeth DiRienzo

A dissertation submitted to the Graduate Faculty of

North Carolina State University

in partial fulfillment of the

requirements for the Degree of

Doctor of Philosophy

Economics

Raleigh

August 2002

Approved By:

_____	_____
Paul Fackler	Barry Goodwin
Co-Chair of Advisory Committee	Co-Chair of Advisory Committee
_____	_____
Ray Palmquist	Jacqueline Hughes-Oliver

Dedication

I dedicate my dissertation to my parents, John and Robin Gallagher and Ivo and Clara DiRienzo, who are truly amazing people.

Biography

Cassandra (Casey) Elizabeth Gallagher was born in Carmel, CA on February 11, 1973 to John P. and Robin M. Gallagher. In 1991, she graduated from Chapel Hill High School in Chapel Hill, NC and enrolled in The Ohio State University. She married Marc E. DiRienzo in 1994 and became Cassandra E. DiRienzo. In 1995, she earned a Bachelor's of Arts degree in Economics. In the fall of 1997, she began her graduate studies in economics at North Carolina State University and, in 2001, she received her Master of Economics.

Acknowledgements

I express my sincere thanks to everyone who has helped me and gave me encouragement during this project. Thanks to my advisors, Paul Fackler and Barry Goodwin for their guidance and insights. I thank my committee members, Ray Palmquist and Jacqueline Hughes-Oliver for their valuable input into my work.

I would like to personally thank Anthony Mancuso and Mary Virginia Atasoy for their friendship. Finally, I would like to thank my husband, Marc DiRienzo, who has been and always will be, my best friend.

Contents

List of Tables vii

List of Figures ix

0.1 U.S. Federal Crop Insurance Program 1

0.1.1 Overview 1

0.1.2 History of the Federal Crop Insurance Corporation 2

0.1.3 Causes of the FCIC's Poor Actuarial Performance 7

0.2 Data 13

0.2.1 Data Transformations and Measures of Relationship 23

0.2.2 Method Used to Remove Time Trend 23

0.2.3 Method Used to Obtain Uniform Variates 29

0.2.4 Summary 36

0.3 Testing the Equality of Cross-Crop Yield Correlation 37

0.3.1 The Sign Test 39

0.3.2 The Friedman Test 48

0.3.3 Conclusion of the Cross-Crop Yield Correlation Tests 51

0.4 The Rate of Spatial Correlation Decay of Crop Yields 56

0.4.1 The Rate of Spatial Correlation Decay Model 57

0.4.2 Conclusion of the Rate of Spatial Correlation Analysis 79

0.5 The Bivariate Distributions of Crop Yields: The Copula Methodology 81

0.5.1 Application to Corn Yields 89

0.5.2 Application to Corn and Soybean Yields 102

0.5.3 Copula Methodology: Crop Insurance Application 105

0.6 Research Summary and Conclusions 113

0.7 Bibliography 120

0.8 Appendix A: Rate of Spatial Correlation Decay 124

0.9 Appendix B: Bivariate Distributions of Crop Yields 125

0.9.1 Appendix B (Part One) 125

0.9.2 Appendix B (Part Two) 128

List of Tables

1 Chow Test of Means Results 27

2 Chow Test of Variances Results 28

3 Durbin-Watson Test Results 29

4 Transform Simulation Results 35

5 Size and Power of Sign Test (Table) 44

6 Sign Test Results 45

7 Friedman Test Results 51

8 Cross-Crop Correlation (Method 1) 54

9 Cross-Crop Correlation (Method 2) 55

10 Cross-Crop Correlation (Method 3) 55

11 Eq. (2) Estimation Results: Illinois 60

12 Eq. (2) Estimation Results: Indiana 60

13 Eq. (2) Estimation Results: Iowa 60

14 Eq. (2) Estimation Results: North Carolina 61

15 Average Spearman Correlation Over Distance Ranges 66

16 Eq. (3) Estimation Results: Illinois 73

17 Eq. (3) Estimation Results: Indiana 73

18 Eq. (3) Estimation Results: Iowa 74

19 Eq. (3) Estimation Results: North Carolina 74
20 Copula Results: Corn Yield (Absolute) 91
21 Copula Results: Corn Yields (Percentage) 91
22 Copula Results by Distance: Illinois (Absolute) 94
23 Copula Results by Distance: Illinois (Percentage) 94
24 Copula Results by Distance: Indiana (Absolute) 95
25 Copula Results by Distance: Indiana (Percentage) 95
26 Copula Results by Distance: Iowa (Absolute) 96
27 Copula Results by Distance: Iowa (Percentage) 96
28 Copula Results by Distance: North Carolina (Absolute) 97
29 Copula Results by Distance: North Carolina (Percentage) 97
30 "Best" to "Second-Best" Odds Ratios: Corn Yields 98
31 "Best" to Normal Odds Ratios: Corn Yields 99
32 Odds Ratio Simulation Results 101
33 Copula Results: Corn and Soybean Yields (Absolute) 102
34 Copula Results: Corn and Soybean Yields (Percentage) 102
35 "Best" to "Second-Best" Odds: Corn and Soybean Yields 103
36 "Best" to Normal Odds: Corn and Soybean Yields 104
37 Crop Insurance Policy Parameters 106
38 Beta($\boldsymbol{p}, \boldsymbol{q}$): Story County Corn and Soybeans 106
39 Probability of Indemnity: Story County Corn and Soybeans 109
40 Indemnity Payments 110

List of Figures

1 Illinois: Mean of Corn Yields By County 15

2 Illinois: Standard Deviation of Corn Yields By County 15

3 Illinois: Mean of Soybean Yields By County 16

4 Illinois: Standard Deviation of Soybean Yields By County 16

5 Indiana: Mean of Corn Yields By County 17

6 Indiana: Standard Deviation of Corn Yields By County 17

7 Indiana: Mean of Soybean Yields By County 18

8 Indiana: Standard Deviation of Soybean Yields By County 18

9 Iowa: Mean of Corn Yields By County 19

10 Iowa: Standard Deviation of Corn Yields By County 19

11 Iowa: Mean of Soybean Yields By County 20

12 Iowa: Standard Deviation of Soybean Yields By County 20

13 North Carolina: Mean of Corn Yields By County 21

14 North Carolina: Standard Deviation of Corn Yields By County 21

15 North Carolina: Mean of Soybean Yields By County 22

16 North Carolina: Standard Deviation of Soybean Yields By County . . 22

17 Power and Size of the Sign Test (Graph) 43

18 Illinois: Corn and Soybean Correlation by County 46

19 Indiana: Corn and Soybean Correlation by County 46
20 Iowa: Corn and Soybean Correlation by County 47
21 North Carolina: Corn and Soybean Correlation by County 47
22 Rate of Spatial Correlation Decay Model: Illinois 61
23 Rate of Spatial Correlation Decay Model: Indiana 62
24 Rate of Spatial Correlation Decay Model: Iowa 62
25 Rate of Spatial Correlation Decay Model: North Carolina 63
26 Illinois: Corn Yield Correlations: Livingston County 68
27 Illinois: Corn Yield Correlations: Shelby County 68
28 Indiana: Corn Yield Correlations: Noble County 69
29 Indiana: Corn Yield Correlations: Marion County 69
30 Iowa: Corn Yield Correlations: Cass County 70
31 Iowa: Corn Yield Correlations: Hamilton County 70
32 North Carolina: Corn Yield Correlations: Chatham County 71
33 North Carolina: Corn Yield Correlations: Sampson County 71
34 Spatial Correlation Decay: Distance and Direction (Iowa) 75
35 Spatial Correlation Decay: Distance and Direction (North Carolina) . 76
36 Ali-Mikhail-Haq Copula 83
37 Cook-Johnson Copula 84
38 Frank Copula 84
39 Gumbel-Hougaard Copula 85
40 Joe Copula 85
41 Normal Copula 86

42 Plackett Copula 86

43 Representative Copula Distribution 108

0.1 U.S. Federal Crop Insurance Program

0.1.1 Overview

Farming is an inherently risky activity as farmers must make production decisions under the uncertainty of such risk factors as weather, potential pest infestations, and price fluctuations. Since the 1930's farmers have been able to transfer part of the production risk, and more recently revenue risk, to the federal government through the U.S. federal crop insurance programs. The objective of the federal crop insurance programs is to alleviate some of the risk that farmers face by providing protection against production and revenue losses. Although there are many different types of insurance programs offered, the basic premise is the same for all programs; covered farmers receive an indemnity payment if their production or revenue falls below a guaranteed level.

The federal crop insurance programs are under the management of the Federal Crop Insurance Corporation (FCIC)[1]. The FCIC was originated largely due to political pressures during the Roosevelt administration under Title V of the Agriculture and Adjustment Act and it has been plagued with financial losses almost every year it has been in existence. The FCIC offers two main types of insurance; traditional yield crop insurance and revenue insurance. Traditional crop insurance has been offered since the FCIC's establishment. It is a multiple-peril insurance program that protects farmers against losses in their level of crop production[2]. In the mid 1990's

[1]The FCIC is a government corporation under the management of the Risk Management Agency, a division of the USDA.

[2]Traditional crop insurance is a multiple-peril insurance policy, meaning that it covers crop yield loss as a result of any natural loss, including unavoidable losses from drought, excessive rain, hail, fire and storm damage.

the FCIC introduced revenue insurance programs. The revenue insurance programs protect farmers against losses in revenue as a result of declines in crop prices, crop yields, or both. The premiums of both types of crop insurance are subsidized by the federal government. The following provides a history of the FCIC and the crop insurance programs supported by the FCIC. In addition, a summary of the research that has explored the causes of the FCIC's poor actuarial performance is provided.

0.1.2 History of the Federal Crop Insurance Corporation

The FCIC was established in 1938 to offer traditional crop insurance to farmers for a limited range of crops in selected areas of the U.S. From 1938 to 1944, traditional crop insurance was available for three types of crops; wheat, cotton, and flax. In each of these years, the premiums collected by the FCIC were less than the indemnities paid. The average loss ratio[3] for this period was equal to 1.65 and the annual subsidies averaged $11.7 million[4]. The program was discontinued in 1944 due to substantial financial loss, but was reestablished in 1945 in response to political pressures. (Goodwin and Smith[10])

From 1945 to 1973, the FCIC expanded by using experimental insurance programs to cover more crops such as corn, oats, barley, rye, tobacco, rice, peanuts, forest products, and citrus fruits. The additional crop coverage increased the av-

[3]The (subsidy-adjusted, see Footnote 4) average loss ratio is equal to:

$$ALR = \frac{1}{N}\Sigma_{n=1}^{N}\frac{I_n}{P_n}$$

where ALR is the average loss ratio, I is the annual indemnity paid for policy n, P is the annual premium paid by policy n, and t = 1...N.

[4]All the figures noted in this paper are adjusted for premium subsidies. The FCIC pays a portion of each farmer's insurance premium in the form of a subsidy. The loss ratios presented in this paper are subsidy-adjusted such that the premiums used in the calculations are only the premiums paid by the farmers and not the combined premium payments made by the FCIC and the farmers.

erage loss ratio to 1.95 from 1945 to 1946 and the annual subsidies grew to $21 million. Although still operating at a loss, the FCIC's average loss ratio decreased to 1.6 from 1947 to 1955. From 1956 to 1973, the FCIC focused on increasing farmer participation by providing insurance for additional crops and by lowering the eligibility requirements which allowed more counties to participate in the program. The FCIC believed that increasing the crop insurance participation rate would lower the average loss ratios. Although participation did not grow extensively during this period, the average loss ratios did fall to 0.9 in eight of the eleven years. To further encourage participation, in 1965 the FCIC lowered the insurance premium rates on most crops. As a result of the lower premiums, the participation rates increased, but so did the indemnity payments and the average loss ratios. In 1970, the Nixon administration canceled the high-loss crop insurance programs[5] and modified other programs to alleviate losses. These changes were successful in reducing the FCIC's average loss ratios to 0.6 for 1971 through 1974. (Goodwin and Smith[10])

Disaster relief payments were established under the 1974 Farm Bill and continued until 1980. During this period, farmers had access to both crop insurance and disaster relief payments. Disaster relief was essentially a free crop insurance program for catastrophic losses[6], as indemnity payments did not require a premium. The disaster relief program and the crop insurance programs were a substantial cost to the federal government. During the period in which both programs were available to farmers, disaster relief payments totaled $3.39 billion and subsidies of premiums totaled $214

[5]The high-loss crop insurance programs were those programs providing production insurance in marginal farmlands and, as a result, were suffering the greatest losses.

[6]Of course, disaster relief payments are *ad hoc* and, therefore, are not necessarily guaranteed.

million.

The 1980 Federal Crop Insurance Act introduced significant changes in the federal crop insurance program. Perhaps the most significant change was the elimination of disaster relief. Additionally, farmers could chose between three levels of traditional crop insurance coverage relative to their average yields. From 1980 to 1984, a farmer's premium was based on an average crop yield for the geographical area in which the farmer was producing. In 1985, the FCIC changed this procedure by introducing the Annual Production History (APH) method for determining a farmer's premium. The APH method bases the farmer's premium on the average of the farmer's actual production over the preceding four to ten years. Further, the 1980 legislation enlisted private insurance companies to sell, service, and share the underwriting risk of the federal crop insurance policies. In accordance with this legislation, the U.S. Department of Agriculture (USDA) reimburses the private insurance companies for the administrative costs associated with selling and servicing the crop insurance policies, including the expenses associated with adjusting claims[7].

In 1994, the Federal Crop Insurance Reform and Department of Agriculture Reorganization Act provided the next major restructuring of the federal crop insurance program by producing two major changes to the program. First, farmers could chose from two basic levels of traditional crop insurance; catastrophic and buy-up. Catastrophic crop insurance is designed to replace the disaster relief by providing farmers with protection against extreme crop losses. Buy-up insurance provides farmers with

[7]The federal government reimburses private insurers for their costs by paying a commission for each policy sold. Further, the FCIC assumes the largest share of the risk associated with underwriting the federal crop insurance policies through the Standard Reinsurance Act (SRA).

protection against smaller, more likely crop losses. Secondly, the legislation required farmers wanting to participate in other USDA programs to purchase a minimum amount of crop insurance.

Two years later, the Federal Agricultural Improvement and Reform Act of 1996 resulted in additional changes to the federal crop insurance program. Prior to the 1996 Act, the USDA offered deficiency programs for several major crops[8]. The deficiency programs offered farmers protection explicitly for loss due to declines in crop prices rather than production losses covered by traditional crop insurance. The deficiency program's indemnity payments were based on a complicated array of pricing mechanisms. The 1996 legislation replaced these deficiency programs with "production flexibility contracts". The production flexibility contracts encouraged farmers to respond to market forces rather than to indemnity payments. These contracts provided farmers with fixed, but declining annual payments that are not associated with crop market prices, thus encouraging farmers to focus their attention on market prices. (GAO Report[5])

From 1996 to 1997 the USDA introduced three new risk management programs for farmers; the Crop Revenue Coverage program, the Income Protection program, and the Revenue Assurance program. These new programs offer farmers protection against revenue losses as a result of a decline in crop prices, crop yields, or both. The revenue insurance programs account for approximately one-third of the total crop insurance sales in the areas where they are offered. The Crop Revenue Coverage program is the largest of the three in terms of participation[9]. The three revenue

[8]These crops included wheat, feed grains, cotton, and rice.

[9]The Crop Revenue Coverage program is also offered in more areas and has more eligible crops

programs differ in the revenue guarantee they provide to farmers. Originally, the Revenue Assurance and Income Protection programs set the revenue level that is to be protected at the time that the crops are being planted. The Revenue Assurance program now allows the option for the protected revenue level to be determined at harvest prices. The Crop Revenue Coverage program determines the protected revenue level at either planting or at harvest, depending on when crop prices are higher.

From 1980 to 1998, the FCIC increased the availability of crop insurance from 30 to 67 crops and from approximately one-half of the U.S. counties to virtually all counties. Participation increased from about ten percent of eligible acres in 1980 to approximately 40 percent in the early 1990s. The requirement introduced by the 1994 Reform Act for farmers to purchase a minimum amount of crop insurance if they wanted to participate in other USDA farm programs, increased participation to over 70 percent of eligible acres in 1994. Despite the numerous revisions to the federal crop insurance programs, the increased number of eligible crops and counties, and the increase in the program's participation rate, the FCIC continues to exhibit poor actuarial performance. From 1990 to 1997 federal costs averaged $1.1 billion annually, totaling over $8.9 billion during this period[10]. (GAO Report[6])

than the other two revenue insurance programs.

[10]However, in recent years, the loss ratio is approximately equal to 1 *if* the loss ratio is not adjusted for premium subsidies such that the premium used in the loss ratio calculation is the combination of the premium subsidy paid by the FCIC and the premium paid by the farmers.

0.1.3 Causes of the FCIC's Poor Actuarial Performance

The reasons offered by the literature for the FCIC's poor actuarial performance have focused on the contractual problems of adverse selection and moral hazard. More recently, systemic risk, which arises from the spatial correlation of crop yields, has been discussed as an additional problem facing crop insurers, although it is not necessarily a factor in the FCIC's poor actuarial performance. An extensive amount of research has examined the contractual problems faced by crop insurers, while relatively little research has explored the spatial correlation of yields. The following provides a summary of these research results.

Contractual Problems

The literature has focused on the contractual problems of adverse selection and moral hazard as causes of the poor actuarial performance of the FCIC. Miranda[17] asserts that adverse selection is the most serious of the contractual problems. Adverse selection occurs when farmers are more informed about the distribution of their own crop yields and are therefore better able to assess the actuarial fairness of the insurance premium than the insurer. Under these circumstances, farmers can more accurately compare their premium with their expected indemnity. Farmers who recognize that their expected indemnity is greater than their premium will be more likely to purchase crop insurance than those farmers whose premium is higher than their expected indemnity. As a result of this asymmetric information, the insurance pool becomes adversely selected. In the long run the insurer will suffer financial losses as the expected indemnity payments exceed the premium payments. If the

insurer were to increase the premium rates, this would result in a smaller and more adversely selected insurance pool. (Miranda[17] and Goodwin[7])

The second contractual problem discussed in the literature is moral hazard. Moral hazard is primarily due to asymmetric information and hidden action. In terms of crop insurance, moral hazard occurs when farmers, after purchasing insurance, adjust their production practices in a manner that increases their chances of collecting an indemnity. The production adjustments made in crop production typically translate into a reduction in the amount of fertilizer or other agricultural chemicals applied to crops. (Goodwin and Smith[10]) This occurs because the FCIC cannot monitor the individual farmer's actions. Lambert[16] and Rogerson[22] argue that farmers should not be expected to practice moral hazard in a particular period because doing so will result in inadequate insurance in future periods[11]. However Vercammen and van Kooten[27], Goodwin and Smith[10], and Goodwin[8] all provide strong empirical evidence supporting the existence of moral hazard in the federal crop insurance program.

Although there remains a debate in the literature, the majority of the literature supports the theory that the asymmetric information in the crop insurance program encourages those farmers who are more likely to receive an indemnity payment to participate in the program. And, once these farmers have purchased insurance, the inability of the FCIC to monitor the farmer's actions encourages farmers to practice moral hazard by reducing the amount of fertilizer and other agricultural chemicals

[11]If the amount of insurance available depends on the farmer's average yield, practicing moral hazard will result in lower yields and a lower average yield, which reduces the amount of insurance available in future periods.

applied to their crops. Clearly, these contractual problems are likely contributors to the FCIC's poor actuarial performance.

Systemic Risk

More recently, the literature has focused on the systemic risk faced by crop insurers as another problem facing the FCIC. Systemic risk, in regard to agricultural production, is primarily an issue of the spatial correlation of crop yields stemming from spatially correlated weather and soil patterns. Crop insurance policies and, therefore, indemnity payments become correlated due to the spatially correlated crop yields. In other words, if a crop insurer has issued several crop insurance policies which insure crop yields in a particular region and these crop yields are spatially correlated, the crop insurer is insuring dependent events. Further, insuring more crop yields in a region in which the crop yields are correlated does not reduce the risk in the insurance pool.

Miranda and Glauber[18] state "...the lack of stochastic independence among individual crop yields defeats insurer efforts to pool crop loss across farms, causing crop insurers to bear a substantially higher risk per unit of premium..." Miranda and Glauber develop a stochastic simulation model of farm-level crop insurance indemnities to measure the systemic risk faced by U.S. crop insurers. They find that crop insurers face portfolio risks that are ten times larger than those faced by private insurers offering more conventional lines of insurance[12]. They argue that the greater portfolio risks faced by crop insurers are due to the correlated indemnity payments

[12]Miranda and Glauber[18] compared crop insurers' portfolio risk to the portfolio risk faced by private automobile and multiple-peril homeowners insurers, among others.

which are a result of spatially correlated yields. Miranda and Glauber find that crop insurers face portfolio risks that are 22 to 49 times greater than if the indemnity payments were independent.

Despite the literature's recent focus on the systemic risk faced by crop insurers, little research has examined the actual spatial dependence structure of crop yields. In theory, there is a multivariate joint distribution of all crops across space that is a function of time, space, and crop. This distribution would provide the necessary information to determine how the spatial dependence structure of crop yields has impacted the crop insurance programs. Further, this distribution would aid crop insurers in developing more actuarially fair crop insurance programs. Although it is not feasible to determine the exact form of this distribution, each of the research topics in this paper explore characteristics of this unknown distribution by making simplifying assumptions.

Conceptually, the parameters of this joint distribution drift at some rate across space. This paper focuses on the drift of three of these parameters. The overall purpose of each of the analyses performed in this paper is to determine the speed at which a parameter drifts, and for some analyses, to determine if the parameter drifts slowly enough to be considered the same over a particular geographical area.

In each of these analyses, U.S. states are considered different geographical areas. State boundaries are used as a method for subdividing and grouping the county-level data[13] such that a state boundary defines a geographical area to be considered. It is important to note that state boundaries were chosen for ease of discussion and report-

[13]See the next section "Data" for a discussion of the county-level data used in these analyses.

ing purposes only and that other geographical areas could be considered. Therefore, the research performed in this paper should not be interpreted as suggesting that the parameters of the joint distribution fundamentally change (and therefore need to be analyzed and modeled differently) as they cross over state boundaries.

The first analysis explores the county-level, cross-crop spatial correlation of yields, which is defined as the correlation between crop yield m and crop yield n, where m and n are different crop yields in the same county. The focus of this first analysis is to determine if the cross-crop correlation parameter drifts slowly enough across space such that it can be considered the same over a particular geographical area, in this a case, a state. If the cross-crop correlation parameter is found not to be the same over a particular region, then the parameter would be considered to be dependent on the location in which it is calculated, or "location-dependent".

The second analysis explores the rate of crop yield spatial correlation decay. The focus of this analysis is to determine the rate at which the correlation between same crop yields drifts or decays as the distance or space between the yields increases. More generally, this analysis explores the drift rate of one of the parameters across space.

The final analysis employs the copula methodology to determine the bivariate copula distribution that "best" describes the spatial dependence structure of crop yields, allowing for an exploration of the higher-order spatial relationships of crop yields. After the "best" bivariate copula distribution is determined, the likelihood value from that estimation procedure is compared to the likelihood value from the Normal bivariate copula estimation using the same crop yield data to gauge the

difference between the likelihood values. More generally, this analysis seeks to determine if the parameters of the bivariate joint distribution of crop yields change slowly enough over a particular geographical region that the bivariate distribution describing crop yields can be considered the same over that region. In particular, this analysis considers if it is appropriate to use the Normal bivariate copula as the bivariate distribution describing crop yields over a particular geographical area. The benefits of the Normal copula are discussed in this analysis section.

Each of these analyses and their corresponding implications for crop insurance are found in the sections 0.3, 0.4, and 0.5 which follow the Data section (0.2).

0.2 Data

This section discusses the data set used in the three analysis sections of this paper. The data set contains the yearly county average harvested corn yields for Illinois, Indiana, Iowa, and North Carolina from 1950 through 2001 and the county average harvested soybean yields for Illinois, Indiana, and Iowa from 1950 through 2001 and for North Carolina from 1956 to 2001[14]. There are 102 counties in Illinois, 92 counties in Indiana, 99 counties in Iowa, and 100 counties in North Carolina. Therefore, the Illinois state data set consists of 102 counties each containing 51 observations of corn and soybean yields; the Indiana state data set consists of 92 counties each containing 51 observations of corn and soybean yields; the Iowa state data set consists of 99 counties each containing 51 observations of corn and soybean yields; and the North Carolina data set consists of 100 counties each containing 51 observations of corn yields and 45 observations of soybean yields. The county average harvested crop yields were obtained from the National Agricultural Statistics Service (NASS), a division of the USDA.

The varieties of crops differed prior to 1950 compared to the more recent varieties, and as a result, the crop yield data prior to 1950 is not necessarily relevant to more current farming practices. Therefore, only the county yield data from 1950 to 2001 is used. The county data from Illinois, Indiana, and Iowa were selected as they are Corn Belt states and have been the focus of much of the current literature. Further, the crop insurance programs discussed in the previous section are available in these three states. The county data from North Carolina was chosen for comparison purposes.

[14]County average harvested soybean yields are not available from 1950 to 1955 in North Carolina.

North Carolina is not a Corn Belt state and the growing conditions are significantly different from those found in Illinois, Indiana, and Iowa. North Carolina is included to explore how the analyses performed in this section can possibly vary given growing conditions different than those found in the Corn Belt states.

The following figures provide data summary statistics in graphical from. For each state and each crop, a color-coded map is included that shows the average and standard deviation of the crop yield[15] by county. The color maps use the RGB (red, green, blue) system such that the county containing the greatest statistic value in its state is colored in the red and the county containing the smallest statistic value is colored in blue. In most of the color maps, the summary statistics show a spatial pattern, in particular in the average crop yield maps. These spatial patterns are explored in considerable detail in the following analysis sections.

[15]The average and standard deviation of the crop yield is over the period 1950 to 2001, with the exception of the North Carolina soybean yields which are over the period 1956 to 2001.

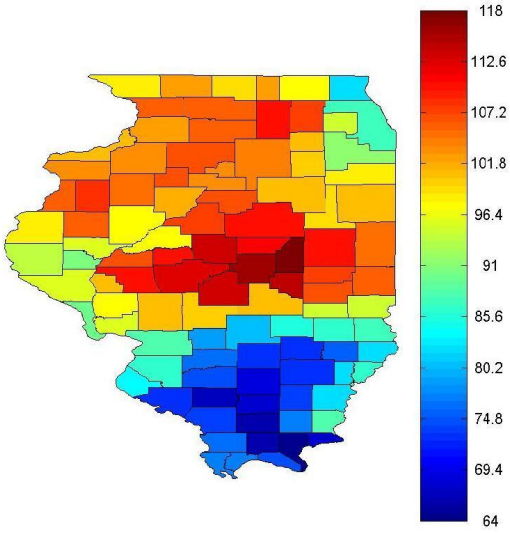


Figure 1: Illinois: Mean of Corn Yields By County

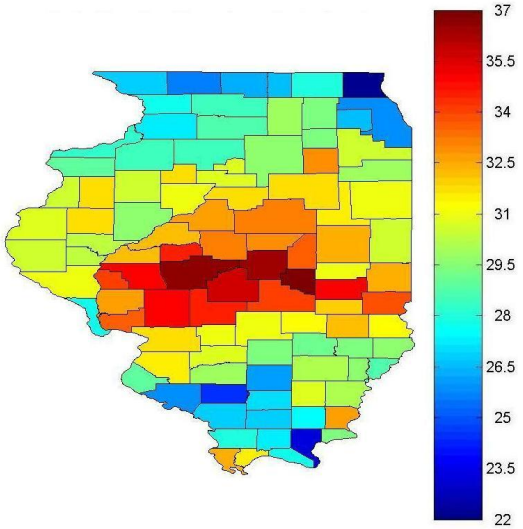


Figure 2: Illinois: Standard Deviation of Corn Yields By County

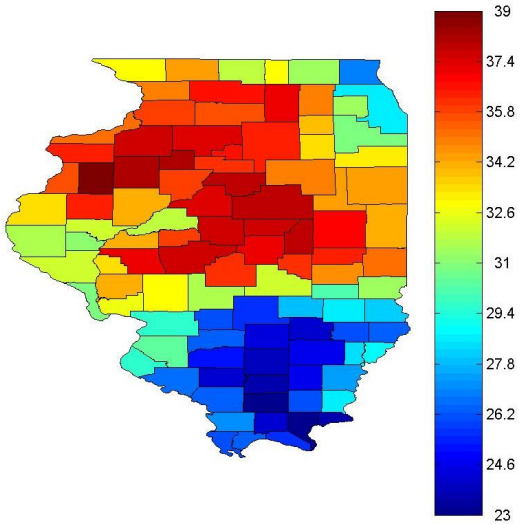


Figure 3: Illinois: Mean of Soybean Yields By County

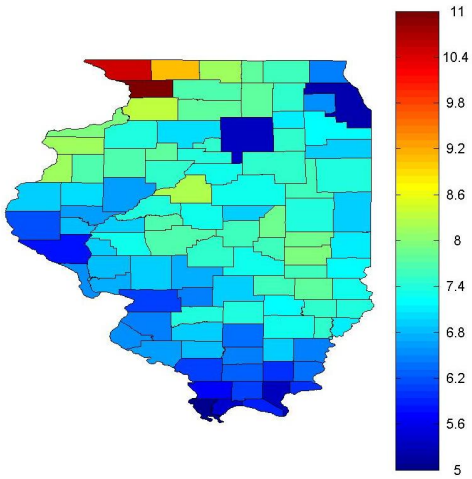


Figure 4: Illinois: Standard Deviation of Soybean Yields By County

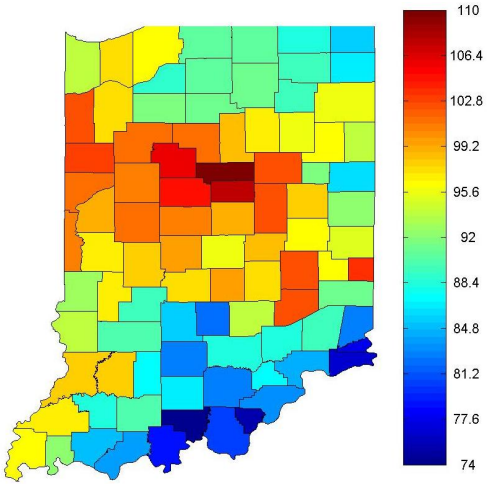


Figure 5: Indiana: Mean of Corn Yields By County

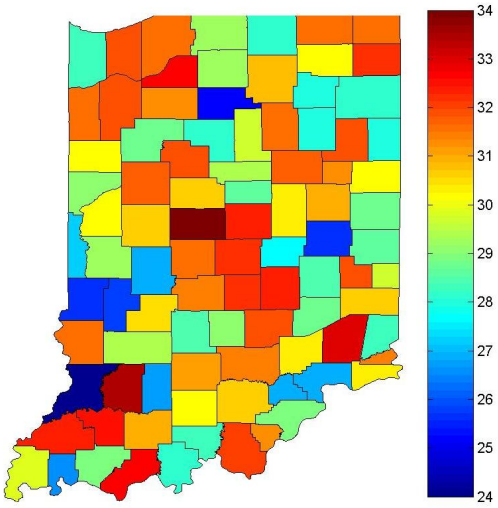


Figure 6: Indiana: Standard Deviation of Corn Yields By County

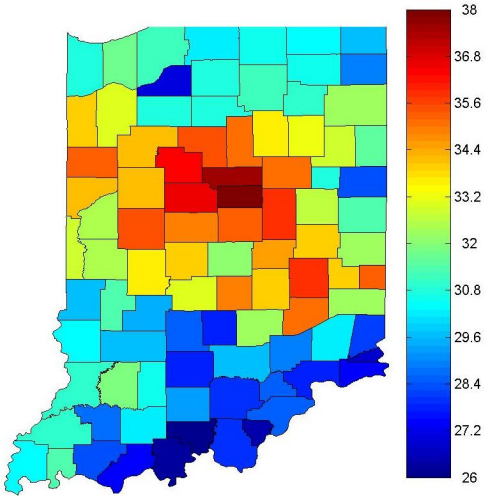


Figure 7: Indiana: Mean of Soybean Yields By County

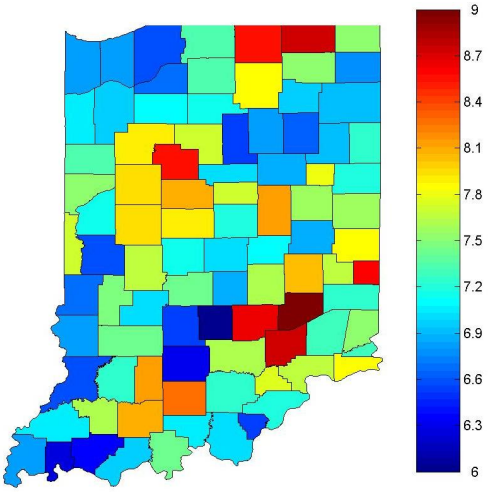


Figure 8: Indiana: Standard Deviation of Soybean Yields By County

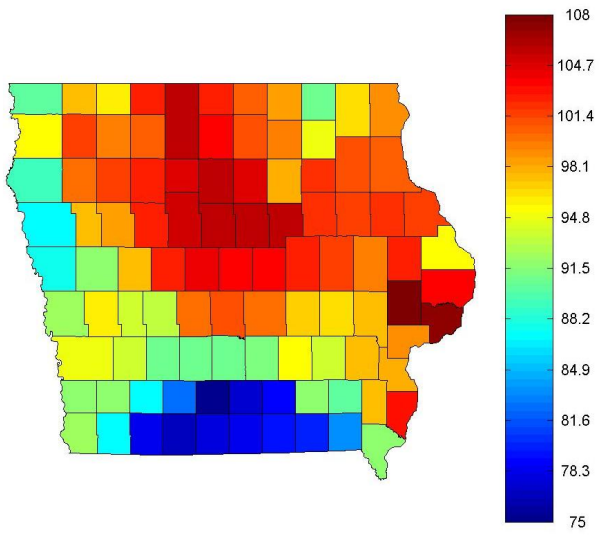


Figure 9: Iowa: Mean of Corn Yields By County

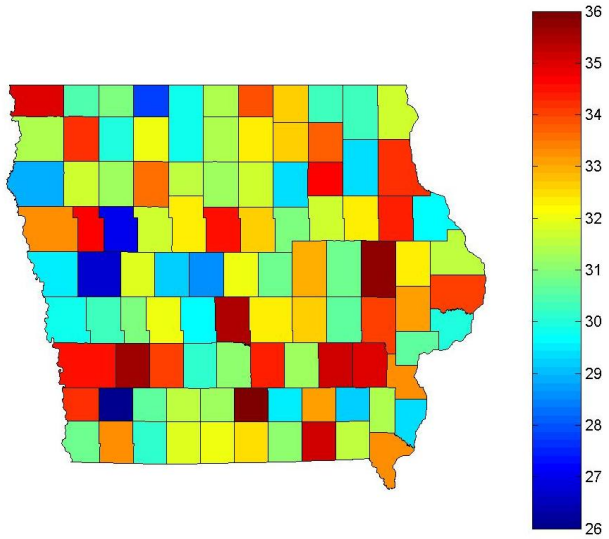


Figure 10: Iowa: Standard Deviation of Corn Yields By County

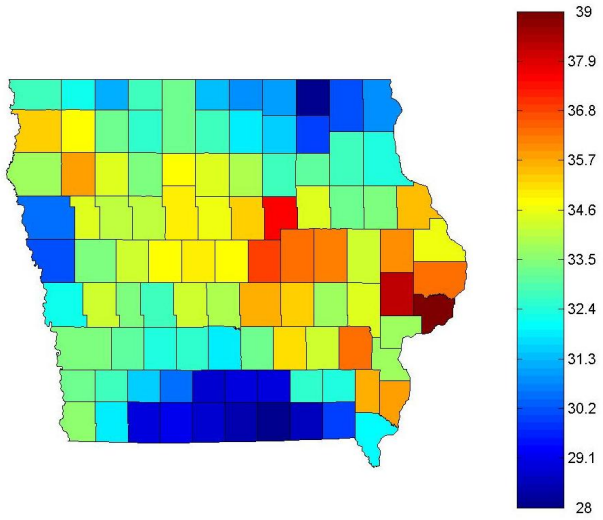


Figure 11: Iowa: Mean of Soybean Yields By County

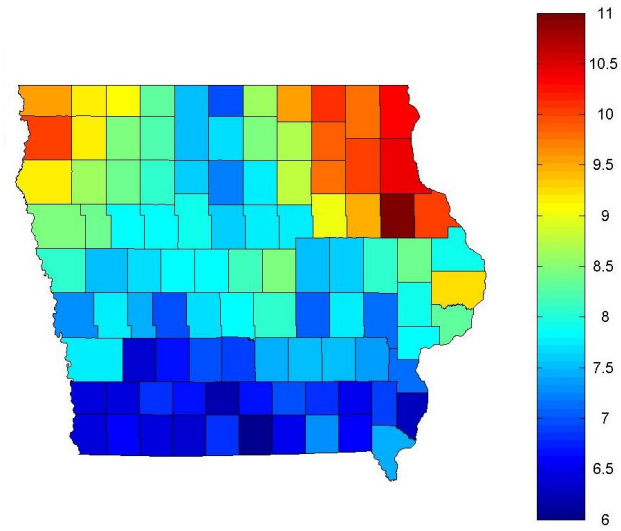


Figure 12: Iowa: Standard Deviation of Soybean Yields By County

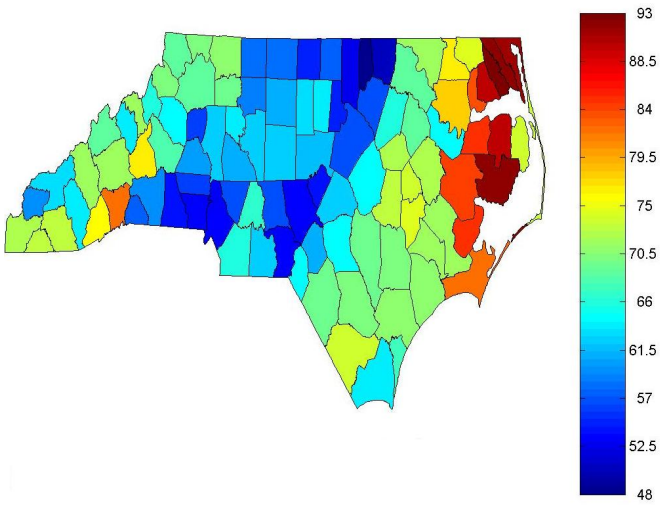


Figure 13: North Carolina: Mean of Corn Yields By County

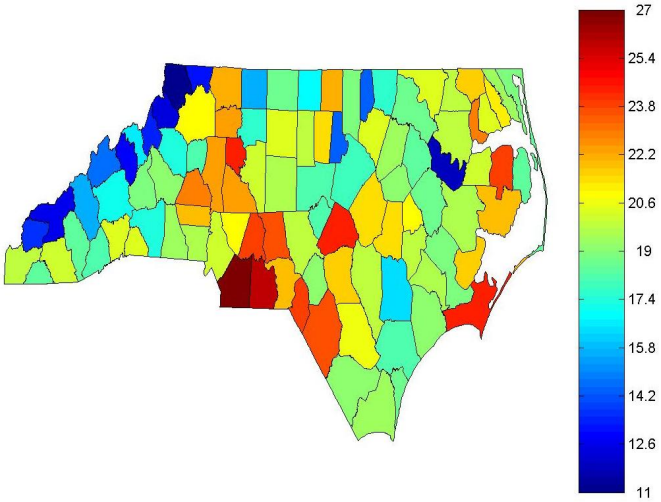


Figure 14: North Carolina: Standard Deviation of Corn Yields By County

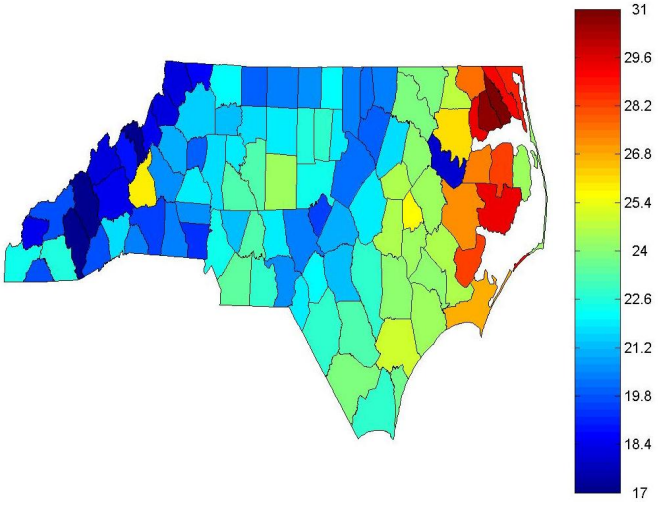


Figure 15: North Carolina: Mean of Soybean Yields By County

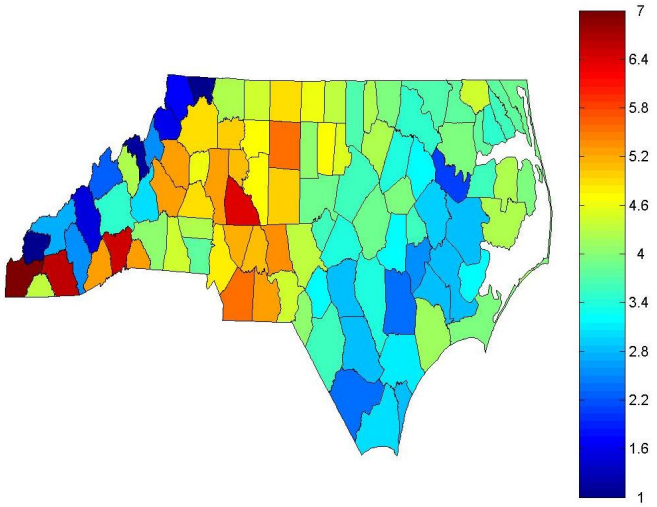


Figure 16: North Carolina: Standard Deviation of Soybean Yields By County

0.2.1 Data Transformations and Measures of Relationship

Two transformations are made to the corn and soybean yield data. The first transform removes the time trend in the data and the second transform converts the detrended data to sets of random, uniform variates by estimating the cumulative density functions (CDFs) of the detrended data. These two transforms are discussed in the following sections. The sets of uniform variates are used in each of the analyses either to estimate the Spearman rank correlation coefficient, or to estimate bivariate copula distributions. The Spearman rank correlation coefficient and copula distributions are both "marginal invariant" measures of relationship, meaning that both measures of relationship are independent to monotone transformations to the variables. The Spearman rank correlation coefficient is discussed in Section 0.3 and copula distributions are discussed in Section 0.5 of this paper.

The following notation will be used to denote observations throughout this paper. The data is triple-indexed $X_{l,c,t}$, where the first subscript l indexes the county or the location of the data, the second subscript c indexes the crop type, and the third subscript t indexes time, or the observation number of crop c in location l, where $t = 1...T$.

0.2.2 Method Used to Remove Time Trend

As discussed in the 1998 GAO report "Problems With New Crop Revenue Insurance Plans"[5], there is no consensus about the method for removing the time trend in crop yields. Goodwin and Ker[9] used a stochastic trend model to capture the time trend. Gallagher[4] removed trend effects by regressing yields and residuals against trend

variables and adjusted the data accordingly. Miranda and Glauber[18] accounted for the time trend by regressing the yields against a second-order polynomial in time. In addition, there is no consensus about how to account for years in which yields were exceptionally low or exceptionally high[16]. Some analyses incorporate all years, but weight exceptionally low and exceptionally high yields relative to the "normal" years, while other analyses discard the extreme yield values entirely. This analysis includes all yield observations and uses the following method to remove the time trend and account for the extreme yield values:

$$\hat{\beta}_{l,c} = (X'X)^{-1}X'y_{l,c}$$

$$\hat{y}_{l,c} = X\hat{\beta}_{l,c}$$

$$\epsilon_{l,c} = \frac{y_{l,c}}{\hat{y}_{l,c}}$$

where $y_{l,c}$ is a (T x 1) vector of crop c yield data from county l, X is a (T x 2) matrix consisting of an intercept and time variable, and $\epsilon_{l,c,t}$ are the elements of the (T x 1) vector $\epsilon_{l,c}$.

This method obtains the OLS parameter estimates[17], but then calculates the residual vector as if the error is multiplicative. Calculating the residual vector using this method accounts for fact that residual values tend to have larger variation as crop yields increase and reduces the residual value for the extreme yield observations. The resulting residual vector should be adjusted for extreme yield observations and not reflect a time trend. This result was tested by using the Durbin-Watson test and variations of the Chow test.

[16]For example, exceedingly low yields were observed in 1988 and 1993, and exceedingly high yields were observed in 1994.

[17]OLS parameter estimates are obtained from the regression $y_{l,c} = X\beta_{l,c} + \epsilon_{l,c}$.

The first test performed was a variation of the Chow test which tests whether the residuals have the same mean in two subsamples. This Chow test, which will be referred to as the "Chow Test of Means", has the following null hypothesis:

$$H_o : \mu^i = \mu^j \ \forall i, j > s$$

where μ^i is the mean of the first subsample which contains the first i $\epsilon_{l,c}$ observations, μ^j is the mean of the second subsample which contains the remaining j $\epsilon_{l,c}$ observations[18], and s is the minimum number of observations required to estimate the regression discussed below[19].

For each residual vector[20], a Chow Test was performed for every pair of subsamples such that the first subsample contains the first i $\epsilon_{l,c}$ observations and the second subsample contains the remaining j $\epsilon_{l,c}$ observations (where $i, j > s$). The restricted regression is defined as:

$$\epsilon_{l,c} = \alpha + \upsilon_{l,c}$$

where α is a vector of constants. The unrestricted regressions are defined as:

$$\epsilon^i_{l,c} = \alpha^i_{l,c} + \upsilon^i_{l,c}$$

$$\epsilon^j_{l,c} = \alpha^j_{l,c} + \upsilon^j_{l,c}$$

where $\epsilon^i_{l,c}$ is an (i x 1) vector containing the first i $\epsilon_{l,c}$ observations and $\epsilon^j_{l,c}$ is an (j x 1) vector containing the remaining j $\epsilon_{l,c}$ observations. As shown in Kennedy[15], an

[18] $j = T - i + 1$, where T is the total number of $\epsilon_{l,c}$ observations.

[19] In this case, $s = 2$.

[20] There are 786 residual vectors in the data set. The corn and soybean yields from the 99 counties in Iowa, 92 counties in Indiana, 102 counties in Illinois, and 100 counties in North Carolina generate 786 residual vectors such that 786 = 2(99) + 2(92) + 2(102) + 2(100).

F statistic for the Chow Test with ν_1 and ν_2 degrees of freedom is formed as:

$$F_{\nu_1,\nu_2} = \frac{RSSE - USSE/\nu_1}{USSE/\nu_2}$$

where RSSE is the sum of the squared errors from the restricted regression, USSE is the sum of the sum of squared errors from the two unrestricted regressions, ν_1 is equal to the number of restrictions being tested, and ν_2 is equal to the number of observations minus the number of regressors (including the intercept) in the unconstrained regression.

The "Chow Test of Means" was applied to each i, j subsample in each of the 786 residual vectors in the data set. The greatest F statistic F_{sup} of each of the 786 residual vectors retained. In other words, for each subsample i, j, a Chow test for each residual vector was performed and the greatest F statistic F_{sup} was retained. Given the 786 residual vectors, this yields 786 F_{sup} statistics. Each of the F_{sup} statistics was compared to the simulated F_{sup} critical value.

As shown by Hansen(96)[12], the asymptotic distribution of F_{sup} is not χ^2. Following Hansen(96) and Hansen(97)[13], the F_{sup} critical value was simulated using the following process:

1. A vector of 51 observations[21] ν was randomly drawn from a standard normal distribution $N(0,1)$.

2. The "Chow Test of Means" was applied to each i, j subsample of ν.

3. The greatest F statistic from the i, j subsample tests F_{sup} was retained.

[21]The vector of 51 observations correspond to a single (51 x 1) residual vector $\epsilon_{l,c}$.

The above process was repeated 25,000 times such that 25,000 F_{sup} values were generated. The $95th$ percentile value of the simulated F_{sup} values was found. This value is the simulated F_{sup} critical value.

The 786 F_{sup} values obtained from the "Chow Test of Means" on the residual data were compared to the simulated F_{sup} critical value. The results are as follows[22]:

Table 1: Chow Test of Means Results

Number of Tests	786
Number of Rejections	37
F_{sup} Critical Value ($\alpha = .05$)	4.53
Percentage of Rejections	4.7%

To explore these results, another simulation was performed in which 786 (51 x 1) residual vectors were drawn from the $N(0, 1)$ distribution 10,000 times. For each of the 10,000 replications, the "Chow Test of Means" was applied to each of the 786 residual vectors such that the F_{sup} value was calculated for each of the 786 residual vectors and compared to the F_{sup} critical value. This process was repeated 10,000 times, and the average percentage of times the "Chow Test of Means" was rejected using 786 residual vectors was 4.8%. The above results indicate that the yield data rejected the "Chow Test of Means" 4.7% of the time. A rejection of 4.7% is an indication that the residuals have the same mean in each of the i, j subsamples.

The second test performed is similar to the "Chow Test of Means" with the exception that it tests whether the residuals have the same variance in two subsamples. This Chow test, which will be referred to as the "Chow Test of Variance", has the

[22]The F critical value for the North Carolina soybean data is 4.41.

following null hypothesis:

$$H_o : \sigma^i = \sigma^j \ \forall i, j > s$$

where σ^i is the variance of the first subsample which contains the first i $\epsilon_{l,c}$ observations, σ^j is the variance of the second subsample which contains the remaining j $\epsilon_{l,c}$ observations[23], and s is the minimum number of observations required to estimate the regression discussed below[24].

The "Chow Test of Variances" was performed using the same procedure as the "Chow Test of Means" with the exception that the i, j subsample means were replaced with the variances. A simulated F_{sup} critical value was calculated using the same procedure as described previously with the exception that the subsample variances replaced the subsample means. The results of the "Chow Test of Variances" are as follows[25]:

Table 2: Chow Test of Variances Results

Number of Tests	786
Number of Rejections	25
F_{sup} Critical Value ($\alpha = .05$)	5.15
Percentage of Rejections	3.2%

To further explore these results, another simulation was performed. This simulation was the same as the simulation performed to explore the "Chow Test of Means" results with the exception that the "Chow Test of Variances" was applied to the residual vectors. The average percentage of times the "Chow Test of Variances" was rejected in the 10,000 replications using 786 residual vectors was 4.9%. The above

[23] $j = T - i + 1$, where T is the total number of $\epsilon_{l,c}$ observations.
[24] In this case, $s = 2$.
[25] The F critical value for the North Carolina soybean data is 5.01.

results indicate that the yield data rejected the "Chow Test of Variances" 3.2% of the time. A rejection of 3.2% of the tests is an indication that the residuals have the same variance in each of the i, j subsamples.

The residuals were also tested for first-order autocorrelation using the Durbin-Watson test. The Durbin-Watson test ($\alpha = .05$) was applied to each of the 786 residual vectors. The results are as follows[26]:

Table 3: Durbin-Watson Test Results

Number of Tests	786
Number of Rejections	28
d_u ($\alpha = .05$)	1.628
d_l ($\alpha = .05$)	1.462
Percentage of Rejections	3.6%

Given an $\alpha = .05$ and the number of tests (786), one would expect to reject approximately 5% of the tests. A rejection of 3.6% of the tests is an indication that the residuals do not exhibit first-order autocorrelation.

It is assumed, given the results of the three tests described above, that the detrending method described previously was successful in removing the time trend from the first two moments of the corn and soybean yield data.

0.2.3 Method Used to Obtain Uniform Variates

The kernel transform with a fixed bandwidth is used throughout this paper as the method for converting the detrended data to uniform variates by estimating the CDF values of the residuals $\epsilon_{l,c,t}$. The purpose of these CDF estimates is to use them in the estimation of the Spearman rank correlation coefficient in Sections 0.3 and 0.4 and

[26]The d_u critical value for the North Carolina soybean data is 1.615 and the d_l critical value for the North Carolina soybean data is 1.430.

in the estimation of bivariate copula distributions in Section 0.5. Spearman's rank correlation coefficient ρ estimates the population product moment correlation[27]:

$$\rho_{mn} = \frac{E\left[(F_m(.) - 0.5)(F_n(.) - 0.5)\right]}{\frac{1}{12}}$$

where ρ_{mn} is the estimate of Spearman's ρ between the random variables m and n, and F_m and F_n are the CDFs of the random variables m and n, respectively[28]. For the purposes of this paper, $F_m(.)$ represents the CDF of crop yield m and $F_n(.)$ represents the CDF of crop yield n. In order to estimate Spearman's ρ, it is necessary to obtain estimates of the CDFs of the crop yields.

Many methods for estimating the CDF values were considered. A description of each follows.

The Rank Transform Method

The rank transform was performed as follows:

1. The yield data was detrended as described in the section "Method Used to Remove Time Trend" and the residual vectors $\epsilon_{l,c}$ were retained.

2. The ranks of the residuals are computed such that the smallest value of $\epsilon_{l,c}$ receives the rank of one, the second smallest value receives the rank of two, and so on. $\tau_{l,c,t}$ represents the rank of the tth residual $\epsilon_{l,c,t}$.

[27]This statistic is commonly estimated from sample data as follows:

$$r_s = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{[n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2]\,[n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2]}}$$

where n is the number of observations, and r_s is the estimate of the Spearman rank correlation coefficient.

[28]The section "Overview of the Sign Test" in this paper and Joag-Dev's[14] paper provide a further discussion of Spearman's ρ.

3. The rank proportion $\pi_{l,c,t}$ of each rank observation was computed as follows:

$$\pi_{l,c,t} = \frac{\tau_{l,c,t}}{T+1}$$

where T is the number of $\epsilon_{l,c}$ observations. The $\pi_{l,c}$ are the CDF estimates such that:

$$\pi_{l,c} = \hat{F}_{\epsilon_{l,c}}(\epsilon_{l,c})$$

The Kernel Transform Method with Fixed Bandwidth

The kernel transform with the optimal fixed bandwidth for a Gaussian kernel was considered. As discussed in Silverman[24], a kernel function satisfies the condition:

$$\int_{-\infty}^{\infty} K(x)dx = 1$$

Usually, K will be a symmetric probability density function such as the normal density. The kernel estimator with kernel K is defined as:

$$\hat{f}(x) = \frac{1}{Th}\Sigma_{t=1}^{T} K\left(\frac{x - X_t}{h}\right)$$

where T is the number of X_t observations, and h is the bandwidth or smoothing parameter.

Choosing the proper bandwidth parameter h is critical in density estimation. Parzen[20] showed that the optimal fixed bandwidth parameter, that is, the parameter that minimizes the mean integrated squared error of the estimate, is given by:

$$h_{opt} = k_2^{-2/5}\left[\int K(x)^2 dx\right]^{1/5}\left[\int f''(y)^2 dy\right]^{-1/5} T^{-1/5}$$

where $k_2 = \int x^2 K(x)dx$, f is the true but unknown density being estimated, T is the number of observations, and f'' represents $\partial^2 f/\partial y^2$. As shown by Silverman[24], if

the unknown density is normal with a variance of σ^2 and a Gaussian kernel is used, the optimal bandwidth reduces to[29]:

$$h_{opt} = 1.06\sigma T^{-1/5}$$

The kernel transform with the optimal fixed bandwidth was performed as follows:

1. The yield data was detrended as described in the section "Method Used to Remove Time Trend" and the residual vectors $\epsilon_{l,c}$ were retained.

2. The standard deviations of the residuals $\sigma_{l,c}$ were calculated and used to compute h_{opt} as described above.

3. The CDF estimates are defined as:

$$\hat{F}_{l,c}(\epsilon) = \frac{1}{Th_{opt}} \sum_{t=1}^{T} \Phi\left(\frac{\epsilon - \epsilon_{l,c,t}}{h_{opt}}\right)$$

where T is the number of $\epsilon_{l,c}$ observations and Φ is the standard Normal CDF.

The Kernel Transform Method with Variable Bandwidth

The kernel transform with a variable bandwidth was considered as well. The kernel estimator with kernel K and a variable bandwidth is defined as:

$$\hat{f}(x) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{hd_{t,k}} K\left(\frac{x - X_t}{hd_{t,k}}\right)$$

where T is the number of X_t observations, $d_{t,k}$ is the distance[30] from X_t to the k^{th} nearest point in the set comprising the other $T-1$ data points, and the product of

[29]This is Silverman's "rule of thumb" estimate.
[30]The distance $d(x, y)$ between two points is defined as $|x - y|$.

h and $d_{t,k}$ is the bandwidth or smoothing parameter[31]. k is typically chosen to be considerably smaller than the sample size such that $k \approx n^{\frac{1}{2}}$.

The kernel transform with variable bandwidth was performed as follows:

1. The yield data was detrended as described in the section "Method Used to Remove Time Trend" and the residual vectors $\epsilon_{l,c}$ were retained.

2. The CDF estimates are defined as:

$$\hat{F}_{l,c}(\epsilon) = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{hd_{t,k}} \Phi\left(\frac{\epsilon - \epsilon_{l,c,t}}{hd_{t,k}}\right)$$

   where Φ is the standard Normal CDF and $k = 7$.

The Transform Method Selected for Analyses

Each of these transformation methods are nonparametric and, as Goodwin and Ker[9] state, "...may offer advantages in capturing the local idiosyncrasies in yield distributions that may not be fully reflected in parameteric specifications." A simulation analysis was performed to determine which of these transformation methods was "best" for the type of data sets used in this paper. The simulation was performed as follows:

1. Two sets of two vectors of 51 observations[32] were randomly drawn from the multivariate normal distribution $\mathrm{N}_{l,mn}(0, \Sigma_{l,mn})$ where $\Sigma_{l,mn} = \begin{bmatrix} 1 & \varrho \\ \varrho & 1 \end{bmatrix}$. Using earlier notation, these two vectors represent $\epsilon_{l,m}$ and $\epsilon_{l,n}$, the detrended yields of crop m and n from location l.

[31] $h = 1.06T^{-1/5}$

[32] The 51 observations correspond to the 51 crop yield observations contained the data set to be used in this analysis.

2. The CDF values of $F_{l,m}(\epsilon_{l,m})$ and $F_{l,n}(\epsilon_{l,n})$ were computed using the Normal CDF $\mathrm{N}(0, \Sigma_{l,mn})$[33] and the CDF values were estimated using the three transform methods described above.

3. The Spearman rank correlation coefficient was calculated four separate times using the four different CDF values obtained in the previous step such that:

$$z_{l,mn,t} = \frac{1}{\sigma_{\hat{F}_{\epsilon_{l,m}}}} \frac{1}{\sigma_{\hat{F}_{\epsilon_{l,n}}}} \left[(\hat{F}_{\epsilon_{l,m}}(\epsilon_{l,m,t}) - \mu_{\hat{F}_{\epsilon_{l,m}}})(\hat{F}_{\epsilon_{l,n}}(\epsilon_{l,n,t}) - \mu_{\hat{F}_{\epsilon_{l,n}}}) \right]$$

where $\sigma_{\hat{F}_{\epsilon_{l,m}}}$ is the standard deviation of $\hat{F}_{\epsilon_{l,m}}$, $\sigma_{\hat{F}_{\epsilon_{l,n}}}$ is the standard deviation of $\hat{F}_{\epsilon_{l,n}}$, $\mu_{\hat{F}_{\epsilon_{l,m}}}$ is the mean of $\hat{F}_{\epsilon_{l,m}}$, and $\mu_{\hat{F}_{\epsilon_{l,n}}}$ is the mean of $\hat{F}_{\epsilon_{l,n}}$. The estimate of Spearman's $\rho_{l,mn}$ is:

$$\hat{\rho}_{l,mn} = \frac{1}{T} \sum_{t=1}^{T} z_{l,mn,t}$$

The estimate of Spearman's ρ that was calculated using the Normal CDF $\mathrm{N}(0, \Sigma_{l,mn})$ will be denoted $\hat{\rho}^{N}_{l,mn}$, the estimate using the rank transform will be denoted $\hat{\rho}^{R}_{l,mn}$, the estimate using the kernel transform with the optimal fixed bandwidth will be denoted $\hat{\rho}^{Kopt}_{l,mn}$, and the estimate using the kernel transform with the variable bandwidth will be denoted $\hat{\rho}^{Kvar}_{l,mn}$.

4. This simulation was repeated 25,000 times for each of the following values of ϱ [0, .1, .2, .3, .4, .5, .6, .7, .8, .9, .99].

5. The mean squared error (MSE)[34] and Root MSE[35] were computed for each of the Spearman $\hat{\rho}_{l,mn}$ estimates that were calculated using the three transform

[33] These values are the true CDF values.
[34] $MSE = Variance + Bias^2$
[35] The Root MSE is the square root of the MSE.

methods described above ($\hat{\rho}^{R}_{l,mn}$, $\hat{\rho}^{Kopt}_{l,mn}$, and $\hat{\rho}^{Kvar}_{l,mn}$)[36].

The Root MSE was used to compare the three estimators of Spearman's ρ. The estimator with the smallest Root MSE represents the "best" estimator, using the Root MSE as the criterion. The results of this analysis are shown in the following table. The table shows the Root MSE for each value of ϱ.

Table 4: Transform Simulation Results

ϱ	$\hat{\rho}^{R}_{l,mn}$	$\hat{\rho}^{Kopt}_{l,mn}$	$\hat{\rho}^{Kvar}_{l,mn}$
0	0.107382	**0.107289**	0.107372
.1	0.111393	**0.111191**	0.111383
.2	0.119002	**0.118719**	0.118972
.3	0.122934	**0.122557**	0.122894
.4	0.129355	**0.128865**	0.129299
.5	0.135311	**0.134726**	0.135222
.6	0.140206	**0.139502**	0.140083
.7	0.145022	**0.144168**	0.144851
.8	0.150780	**0.149816**	0.150597
.9	0.155306	**0.154118**	0.155003
.99	0.162587	**0.161201**	0.162205

The Root MSE for the $\hat{\rho}^{R}_{l,mn}$, $\hat{\rho}^{Kopt}_{l,mn}$, and $\hat{\rho}^{Kvar}_{l,mn}$ estimates increase as ϱ increases. The results indicate that for all values of ϱ, the $\hat{\rho}^{Kopt}_{l,mn}$ estimates have the smallest Root MSE. For this reason, the kernel transform with the fixed optimal bandwidth is chosen as the method to transform the detrended data into sets of uniform variates.

[36]The Spearman ρ estimates that were calculated using the Normal CDF $N(0, \Sigma_{l,mn})$ ($\hat{\rho}^{N}_{l,mn}$) were used to compute the bias of each of the transform methods such that:

$$Bias^2 = (\hat{\rho}^{transformmethod}_{l,mn} - \hat{\rho}^{N}_{l,mn})^2$$

0.2.4 Summary

It is assumed that the detrending method described previously was successful in removing the time trend from the first two moments of the corn and soybean yield data. The kernel transform with a fixed bandwidth described above converted the detrended data to uniform variates such that the transformed data set used in the following analyses are sets of random, uniform variates. The data discussed in the following analyses (unless otherwise stated) refers to this transformed data. Further, the notation used throughout this paper is consistent with the notation described in the section "Data Transformations and Measures of Relationship".

0.3 Testing the Equality of Cross-Crop Yield Correlation

The focus of this analysis is to determine if the county-level, cross-crop yield correlation parameter changes slowly enough across space that it can be considered the same over a particular geographical area, in this case, a state. The county-level cross-crop yield correlation $\rho_{x,mn}$ is the correlation between the crop yield m and crop yield n for county x. If t crops are produced in county x, the cross-crop yield correlation matrix Ω_x for county x is a (t x t) matrix with ones down the diagonal and $\rho_{x,mn}$ is the m,nth element of Ω_x.

Cross-crop yield correlation is of particular interest to whole-farm revenue insurers[37]. Whole-farm revenue insurance offers farmers protection that is based on the "whole farm's" revenue. In other words, an insured farmer will receive an indemnity payment if the total revenue earned by his farm that period falls below a predetermined level. The cross-crop yield correlation between the insured farmer's crop yields affects the portfolio risk faced by the whole-farm revenue insurer. The effect on a whole-farm revenue insurer's portfolio risk can be observed through a simple revenue function. The revenue earned by a farmer in any given period is simply:

$$\pi_f = \Sigma_{i=1}^{t} p_i y_i$$

where π_f is the revenue earned by farmer f, t is the number of crops produced by farmer f, p_i is the price of crop i, and y_i is the yield of crop i. If the yield of crop m is correlated with the yield of crop n, this cross-crop yield correlation becomes a key factor when assessing the portfolio risk faced by the insurer. A strong positive

[37]The FCIC exclusively offers whole-farm revenue insurance.

correlation between crop yield m and n indicates that if a low yield for crop m is observed, it is likely that a low yield for crop n will also be observed. This relationship between the crop yields implies that the insurer's portfolio is poorly diversified as low yields are not off-set by high yields. A strong negative correlation between crop yield m and crop yield n indicates that if a low yield for crop m is observed, it is likely that a high yield for crop n will be observed. This implies that the insurer's portfolio is well-diversified as low yields are counter-balanced by high yields.

In the above discussion of whole-farm revenue insurance, the portfolio risk faced by the whole-farm insurer with one policy is considered. However, in general practice, whole-farm revenue insurer issues insurance to many farmers in many counties. In assessing the portfolio risk faced by the whole-farm insurer and in calculating the insurance premiums for the whole-farm revenue insurance policies, the question arises as to how the cross-crop yield correlation should be calculated. The answer largely depends on whether or not the cross-crop yield correlation is constant across space, or if it depends on location. Alternatively, the answer depends on whether the county-level, cross-crop yield correlation parameter changes slowly enough across space that it can be considered the same over a particular geographical area.

It is widely known that yields of different types of crops are correlated. However, it is unknown if the cross-crop yield correlation is constant across space, or if it depends on the location of the county of interest. In other words, it is unknown if the cross-crop yield correlation between the crop m and n county-level yield data from county x $\rho_{x,mn}$ is the same as the cross-crop yield correlation between the crop m and n county-level yield data from county y $\rho_{y,mn}$.

0.3.1 The Sign Test

Overview of the Sign Test

The nonparametric Sign Test discussed in Conover[1] is used to explore the equality of the cross-crop yield correlations across space. The Sign Test is used in this context to test whether or not the cross-crop yield correlation between crop yield m and n calculated in county x is the same as the cross-crop yield correlation between crop yield m and n in county y, where county x and y are different counties located in the same state. The null hypothesis of the Sign Test is:

$$H_o : \rho_{x,mn} = \rho_{y,mn}$$

where m indexes crop m, n indexes crop n, ρ represents the Spearman rank correlation coefficient, x indexes county x, and y indexes county y, where county x and y are different counties located in the same state.

As discussed in Joag-Dev[14], the Spearman rank correlation coefficient is an estimator of the product moment correlation between $F_m(.)$ and $F_n(.)$:

$$\rho_{mn} = \frac{E\left[(F_m(.) - 0.5)(F_n(.) - 0.5)\right]}{\frac{1}{12}} \tag{1}$$

where $F_m(.)$ is the cumulative density function (CDF) of the random variable m and $F_n(.)$ is the CDF of the random variable n. For this analysis, $F_m(.)$ represents the CDF of crop yield m and $F_n(.)$ is the CDF of crop yield n. The Spearman rank correlation coefficient is chosen as the measure of correlation between crop yields because it is a "marginal invariant" measure of relationship, as discussed in the section "Data Transformations and Measures of Relationship".

The Sign Test Procedure

The Sign Test procedure for a single pair-wise test of the null hypotheses[38] $H_o : \rho_{x,mn} = \rho_{y,mn}$ is described as follows[39]:

1. For each county x and y, the yield data for the crops m and n are detrended using the method described in the section "Method Used to Remove Time Trend".

2. The kernel transform method with fixed bandwidth is used to estimate the CDFs of the crop yields, denoted $\hat{F}_{x,m}(.)$, $\hat{F}_{x,n}(.)$, $\hat{F}_{y,m}(.)$, and $\hat{F}_{y,n}(.)$[40] as described in the section "Method Used to Obtain Uniform Variates".

3. The mean is subtracted from the CDF estimates and then the centered data for each county are multiplied together such that:

$$z_{x,mn} = (\hat{F}_{x,m}(\epsilon_{x,m}) - 0.5)(\hat{F}_{x,n}(\epsilon_{x,n}) - 0.5)$$

$$z_{y,mn} = (\hat{F}_{y,m}(\epsilon_{y,m}) - 0.5)(\hat{F}_{y,n}(\epsilon_{y,n}) - 0.5)$$

where $z_{x,mn}$ and $z_{y,mn}$ are (T x 1) vectors such that the mean of $z_{x,mn}$ is an estimate of $\rho_{x,mn}$ (Eq. (1)) and the mean of $z_{y,mn}$ is an estimate of $\rho_{y,mn}$ (Eq. (1)).

[38]The data is indexed as described in the section"Data Transformations and Measures of Relationship".

[39]Given that there are more than two counties located in state, there are many unique pair-wise tests of the null hypothesis that can be performed in a state. The following discusses a single pair-wise test of the null hypothesis.

[40]The CDFs are estimates of the CDFs $F_m(.)$ and $F_n(.)$ in Eq. (1).

4. The number of times $z_{x,mn,t} > z_{y,mn,t}$ is counted[41].

$$Z_{xy,mn} = \sum_{t=1}^{T} I(z_{x,mn,t} - z_{y,mn,t} > 0)$$

where I is an indicator function that takes on the value one when $z_{x,mn,t} - z_{y,mn,t} > 0$.

5. Under the null hypothesis that the Spearman cross-crop yield correlation between crop m and n in county x is equal to the Spearman cross-crop yield correlation between crop m and n in county y (H_o: $\rho_{x,mn} = \rho_{y,mn}$), $Z_{xy,mn}$ is distributed Binomial(T, 0.5)[42]. The test statistic $Z_{xy,mn}$ is compared to the appropriate critical values from Binomial(T, 0.5).

The Size and Power of the Sign Test

The size and the power of the sign test were explored using a Monte Carlo simulation analysis described as follows:

1. Two sets of two vectors of 51 observations[43] were randomly drawn from the multivariate normal distributions $N_{x,mn}(0,\Sigma_{x,mn})$ and $N_{y,mn}(0,\Sigma_{y,mn})$ where x represents the first set of draws and y represents the second set of draws, $\Sigma_{x,mn} = \begin{bmatrix} 1 & \varrho_x \\ \varrho_x & 1 \end{bmatrix}$, $\Sigma_{y,mn} = \begin{bmatrix} 1 & \varrho_y \\ \varrho_y & 1 \end{bmatrix}$. Using earlier notation, these four vectors represent $\epsilon_{x,m}$, $\epsilon_{x,n}$, $\epsilon_{y,m}$, and $\epsilon_{y,n}$.

2. The kernel transform method with fixed bandwidth CDF estimates of $F_{x,m}(.)$, $F_{x,n}(.)$, $F_{y,m}(.)$, and $F_{y,n}$ were calculated.

[41] Ties ($z_{x,mn,t} = z_{y,mn,t}$) were not found.

[42] Intuitively, if the mean of $z_{x,mn}$ is equal to the mean of $z_{y,mn}$ (if the estimate of $\rho_{x,mn}$ is equal to the estimate of $\rho_{y,mn}$), then one would expect the $z_{x,mn}$ observations to exceed the $z_{y,mn}$ observations approximately one-half or 50% of the time.

[43] The 51 observations correspond to the 51 crop yield observations contained the data set to be used in this analysis.

3. $z_{x,mn}$, $z_{y,mn}$, and $Z_{xy,mn}$ were calculated as described in previous section.

4. The test statistic $Z_{xy,mn}$ was compared to the appropriate critical values from Binomial(T, 0.5) with $\alpha = 0.05$.

This process was repeated 25,000 times for each combination[44] of ϱ_x and ϱ_y values on the range [0, .1, .2, .3, .4, .5, .6, .7, .8, .9, .99] and the number of times the null hypothesis was rejected was retained. For each combination of ϱ_x and ϱ_y values, the percentage of null hypothesis rejections to the number of tests (25,000) was calculated[45]. The results of this Monte Carlo simulation provide an indication of the size and power of the Sign Test. The size and power of the Sign Test presented in graphical form is shown in the following figure. The values of ϱ_x and ϱ_y are on the x and y-axes and the percentage of rejections are on the z-axis.

[44]There are 121 possible combinations and 66 of them are unique.

[45]Percentage of rejections $= 100 * \left(\frac{Number\,of\,Tests\,Rejected}{25{,}000}\right)$

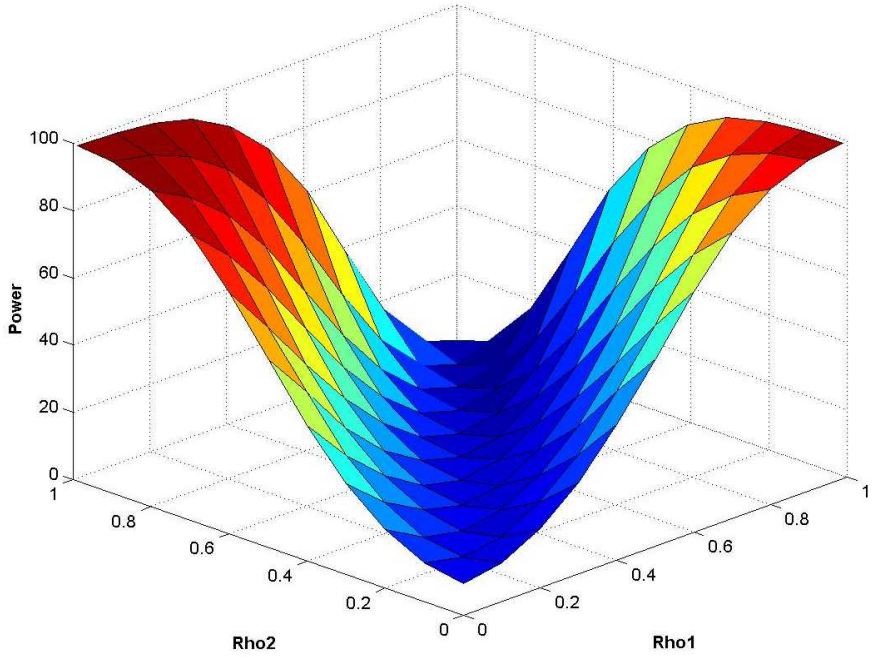


Figure 17: Power and Size of the Sign Test (Graph)

The size and power of the Sign Test presented in table form are as follows.

Table 5: Size and Power of Sign Test (Table)

ϱ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
0	9.5	11.7	17.5	27.0	40.2	55.8	71.6	85.1	94.0	98.6	99.8
0.1		9.2	11.0	17.4	27.2	42.0	57.4	74.5	87.9	96.4	99.3
0.2			8.7	11.0	17.1	27.4	42.1	60.8	78.5	92.0	98.1
0.3				8.3	10.0	16.6	28.1	44.9	64.7	84.0	95.2
0.4					7.6	9.5	16.4	29.0	47.9	70.9	88.9
0.5						6.3	8.1	16.2	30.6	53.7	78.1
0.6							5.1	7.2	16.2	34.5	61.5
0.7								3.8	6.1	16.9	39.4
0.8									2.5	4.9	18.2
0.9										1.2	4.3
0.99											1.0

The simulation results indicate that the Sign Test performs well when the correlation values are in the middle range, but slightly over-rejects when the correlation values are low and slightly under-rejects when the correlation approaches one. Although the Sign Test has difficulty rejecting finer differences of the ϱ values, its power curve shown is steeply sloped.

Sign Test Results

This analysis considered corn and soybean yield data from the four states Illinois, Indiana, Iowa, and North Carolina as described in the "Data" section. The results of the Sign Test performed on the county-level crop yield data from each state are presented below. For each state, the number of pair-wise Sign Tests that were performed is given with the percentage of the Sign Tests that were rejected[46].

[46]Note that in order for a county's yield data to be considered it had to have a minimum of 25 of the 51 possible observations (22 of the possible 45 for North Carolina), or (approximately) 50% of the yield data for both corn and soybeans.

Table 6: Sign Test Results

State	**Percentage of Rejections** ($\alpha = 0.05$)	**Number of Tests**
Illinois	7.8%	5151
Indiana	2.8%	4186
Iowa	3.3%	4851
North Carolina	3.8%	3873

As shown in the above table, the Sign Test is rejected approximately twice as often using the county-level yield data from Illinois compared to the other states considered, indicating that the Spearman correlation coefficient between the corn and soybean yields in Illinois may behave differently than the other states considered in this analysis.

To further explore the cross-crop correlation between corn and soybean yields, the following figures show color maps of the Spearman correlation between corn and soybean yields by county for each state. Again, the color maps use the RGB (red, green, blue) system such that the county containing the greatest statistic value in its state is colored in red and the county containing the smallest statistic value is colored in blue.

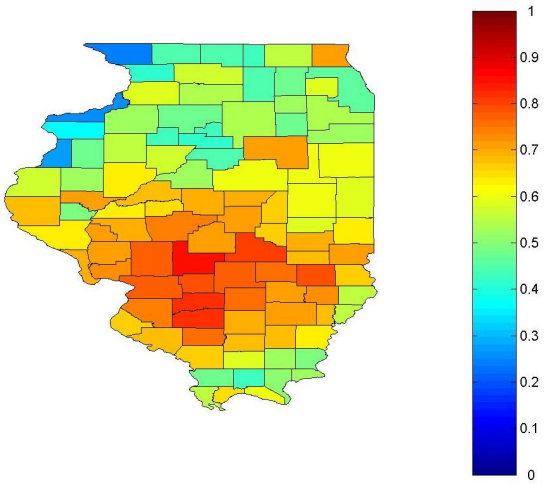


Figure 18: Illinois: Corn and Soybean Correlation by County

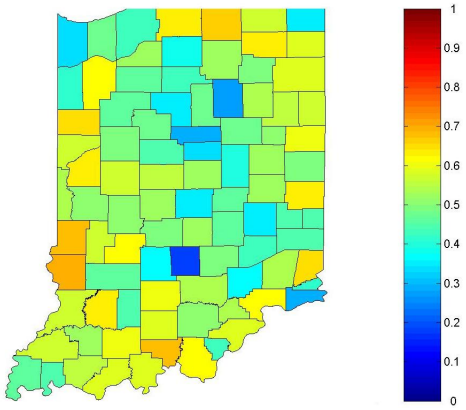


Figure 19: Indiana: Corn and Soybean Correlation by County

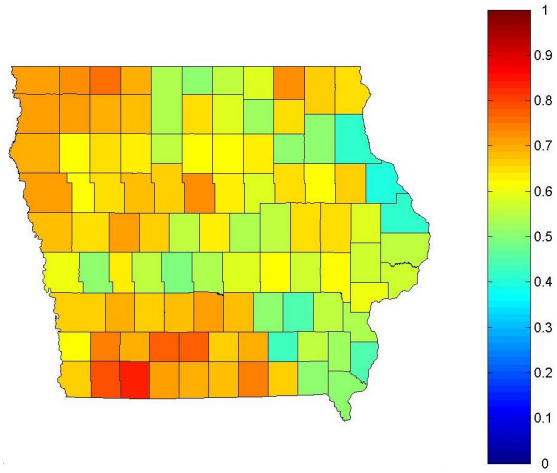


Figure 20: Iowa: Corn and Soybean Correlation by County

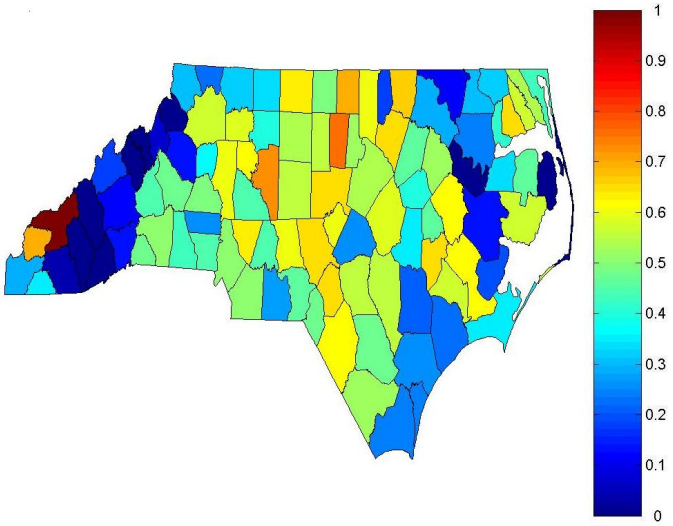


Figure 21: North Carolina: Corn and Soybean Correlation by County

The Illinois map shows the greatest variation in cross-crop correlations as a block of counties beginning in the central region of the state and continuing to the southern half of the state show a greater cross-crop correlation compared to the northern half of the state and the extreme southern counties. These results raise the question of whether or not the lower percentage of Sign Test rejections in Indiana, Iowa, and North Carolina is an indication that $\rho_{x,mn} = \rho_{y,mn}$ for all x and y in these states and that $\rho_{x,mn} \neq \rho_{y,mn}$ for all x and y in Illinois. In other words, can we consider $\rho_{l,mn}$ as constant across all counties in a given state?

As discussed in Conover[1], the Friedman Test is an extension of the Sign Test. The null hypothesis of the Friedman Test in the context of this analysis is:

$$H_o : \rho_{x,mn} = \rho_{y,mn} \; \forall x, y$$

The Friedman test extends the pair-wise null hypothesis of the Sign Test by testing the above joint hypothesis.

0.3.2 The Friedman Test

The Friedman Test is designed to test the null hypothesis that the probability distributions of k "treatments" are identical versus the alternative that at least two of the distributions differ in location. In the context of this analysis, this translates into testing the joint hypothesis that the mean of the $z_{x,mn}$ is equal to the mean of $z_{y,mn}$ for all counties x and y in a given state. Recall that the mean of $z_{l,mn}$ is an estimate of the Spearman rank correlation coefficient $\rho_{l,mn}$.

As discussed in Wackerly, Mendenhall, and Scheaffer[28] the test is based on a statistic that is a rank analogue of Sum of Squared Treatments (SST):

$$SST = b\Sigma_{i=1}^{k}(\bar{Y}_{i\cdot} - \bar{Y})^2 = \Sigma_{i=1}^{k}\frac{Y_{i\cdot}^2}{b} - \frac{1}{bk}(\Sigma_{j=1}^{b}\Sigma_{i=1}^{k}Y_{ij})^2$$

where $\bar{Y}_{i\cdot}$ represents the average for all observations for treatment i, $Y_{i\cdot}^2$ represents the square of the sum of observations for treatment i, k is the number of treatments, and b is the number of blocks, or number of observations for each treatment. In this analysis, the counties in a given state represent the k treatments, the $z_{l,mn}$ observations represent the Y observations, and the number of $z_{l,mn}$ observations is represented by b[47].

The following describes the steps of the Friedman Test applied to the county-level corn and soybean yield data from each of the states in consideration.

1. For each county x and y, the yield data for the crops m and n are detrended using the method described in the section "Method Used to Remove Time Trend".

2. The kernel transform method with fixed bandwidth is used to estimate the CDFs of the crop yields as described in the section "Method Used to Obtain Uniform Variates".

3. The $z_{l,mn}$ vectors are calculated as described previously. For each state, there are L (T x 1) $z_{l,mn}$ vectors which can be combined to form a (L x T) matrix of $z_{l,mn}$ values.

4. The (L x T) matrix of $z_{l,mn}$ values are ranked across time. In other words,

[47]For Illinois, Indiana, and Iowa, b is 51 and for North Carolina, b is 45.

each column of the (L x T) matrix is ranked such that the ranks are on the interval [1:L] and $R(z_{l,mn,t})$ is the rank of the tth observation of $z_{l,mn}$[48].

5. The ranks are summed over county such that $R_l = \sum_{l=1}^{L} R(z_{l,mn,t})$.

6. The statistic $A2$ is calculated[49]:

$$A2 = \frac{TL(L+1)(2L+1)}{6}$$

7. The statistic $B2$ is calculated:

$$B2 = \frac{1}{T}\sum_{l=1}^{L} R_l$$

8. The $T2$ test statistic is calculated:

$$T2 = \frac{(T-1)(B2 - \frac{TL(L+1)^2}{4})}{A2 - B2}$$

The null hypothesis is rejected if $T2$ exceeds the 1 - α quantile of the F distribution with υ_1 = (L - 1) and υ_2 = (L - 1)(T - 1). An $\alpha = 0.05$ was used in this analysis.

The Friedman Test Results

The corn and soybean yield data from four states Illinois, Indiana, Iowa, and North Carolina, as described in the "Data" section, were considered again for the Friedman test. The results of the Friedman Test performed on these crops and states are

[48]The ranking is done such that the smallest value in each column is given the value of one, the second smallest the value of two, and so forth. If two elements in the same column are equal, each element is given the average of the ranks. Therefore, if two elements were tied for the third rank, then they each get a rank of 3.5 = (3+4)/2. There were no ties for the data used in this analysis.

[49]There were no ties in the ranks. As a result, the $A2$ statistic presented here is the simplified form of the statistic that is used when there are no ties in the ranks.

presented below. For each state, the $T2$ test statistic, critical value (for $\alpha = 0.05$), and p-value are given[50].

Table 7: Friedman Test Results

State	**Test Statistic** ($\alpha = 0.05$)	**Critical Value**	***p*-Value**
Illinois	1.4420	1.2454	0.0027
Indiana	0.8425	1.2596	0.8564
Iowa	0.7503	1.2496	0.9682
North Carolina	0.7836	1.2489	0.9434

The results of the Friedman Test indicate that we fail to reject the null hypothesis (H_o: $\rho_{x,mn} = \rho_{y,mn} \forall x, y$) for the Indiana, Iowa, and North Carolina corn and soybean yield data and we reject the null hypothesis for the Illinois corn and soybean yield data. This result coincides with the results from the Sign Test as the Illinois corn and soybean yield data is rejected approximately twice as often compared to the other states. These results indicate that there is a greater difference in the cross-crop yield correlation across the counties in Illinois compared to the other states considered, and by the Friedman Test, we reject the null hypothesis that all the county-level cross-crop yield correlations in Illinois are jointly equal to each other.

0.3.3 Conclusion of the Cross-Crop Yield Correlation Tests

The results of the cross-crop yield correlation analyses indicate that in the state of Illinois, the cross-crop correlation between the county-level corn and soybean yields is location dependent, such that the correlation measure is dependent on the county in which it is calculated. Further, the cross-crop correlation between the county-

[50]Note that in order for a county's yield data to be considered it had to have a minimum of 38 of the 51 possible observations (22 of the possible 45 for North Carolina), or (approximately) 50% of the yield data for both corn and soybeans.

level corn and soybean yields in Indiana, Iowa, and North Carolina was found not to be location dependent, such that the cross-crop correlation between the county-level corn and soybean yields does not depend on the county in which it is calculated. More generally, the cross-crop correlation parameter changes slowly enough over the states Indiana, Iowa, and North Carolina to be considered the same over these respective areas. However, the cross-crop correlation parameter does not change slowly enough over the state of Illinois to be considered the same over this area. These results have the most direct implications for whole-farm revenue insurance.

The results of these analyses indicate that whole-farm revenue insurers need to calculate the cross-crop correlation for Illinois corn and soybean yields by county, as the cross-crop correlation is location or county dependent. However, whole-farm revenue insurers do not need to calculate the cross-crop correlation between corn and soybean yields by county in Indiana, Iowa, and North Carolina, as the cross-crop correlation is not location, or county dependent. This implies that alternate methods of calculating the cross-crop correlation could be used in those states in which the cross-crop correlation is found not to be location dependent. One alternate method of calculating the cross-crop correlation is to pool the county-level data to estimate a single cross-crop correlation parameter for the state. Another method is to estimate the cross-crop correlation parameter for every county in a given state and then calculate a weighted average of the correlation parameters which yields a single cross-crop correlation parameter for the state[51]. Finally, state-level data could be used to estimate the cross-crop correlation parameter. There are many

[51]A weighted average could be calculated by weighting each county-level correlation parameter by the number of crop yield observations used to estimate it.

other methods available to calculate the cross-crop correlation parameter, and which method is optimal depends on the availability of county-level and state-level data, researcher preferences and *a priori* beliefs, among other considerations.

Alternate methods for calculating the cross-crop correlation parameter in states in which the cross-crop correlation is found not to be location dependent can lead to better estimates of the cross-crop correlation. In general, over a given number of years, it is common for a county not to produce a particular crop for some of the years considered. For example, over a period of 50 years, it is likely that any given county did not to produce a given crop in at least one of the years in the 50 year period, resulting in missing county-level yield data. Therefore, methods of calculating the cross-crop correlation that pool county-level data, or use state-level data to calculate a single cross-crop correlation parameter for the state, may lead to better estimates of the cross-crop correlation, as calculating the cross-crop correlation for a particular county relies solely on the availability of that county's data.

As an example, the cross-crop correlation between corn and soybean yields in Indiana, Iowa, and North Carolina were calculated using each of the three methods described above[52]. The first method stacks the detrended corn yield data for each county in a given state into a vector c_s and the detrended soybean yield data for each county in a given state into a vector s_s, where the subscript s denotes the state. The correlation is then calculated between the two vectors c_s and s_s. The following are the estimates of the cross-crop correlation between corn and soybean yields in each of the states considered[53].

[52]The cross-crop correlation was found not to be location or county dependent in these states.

[53]In order for a county's data to be included, it must contain at least 50% of the possible number

Table 8: Cross-Crop Correlation (Method 1)

State	Cross-Crop Correlation Estimate
Indiana	0.5138
Iowa	0.6323
North Carolina	0.3458

The cross-crop correlation was also calculated using the second method described above in which the cross-crop correlation is calculated by county and then a weighted average is calculated from the county-level correlation estimates. A weighted average is calculated by giving the counties with the largest number of yield observations from which the correlation estimate is based the greatest weight and giving the counties with the smallest number of yield observations the least weight. Mathematically, the weighting scheme is defined as:

$$\hat{\rho}_s = \frac{1}{T} \sum_{l=1}^{L} t_l \hat{\rho}_l$$

where $\hat{\rho}_s$ is the weighted average cross-crop correlation for the state s, T is the total number of yield observations from all counties in state s, L is the total number of counties in state s, and t is the number of yield observations in county l. The following are estimates of the cross-crop correlation between corn and soybean yields in each of the states considered[54].

of observations, which is 26 observations in Indiana and Iowa and 23 observations in North Carolina.

[54]In order of a county to be included, it must contain at least 50% of the possible number of observations, which is 26 observations in Indiana and Iowa and 23 observations in North Carolina.

Table 9: Cross-Crop Correlation (Method 2)

State	Cross-Crop Correlation Estimate
Indiana	0.5070
Iowa	0.6238
North Carolina	0.3430

Finally, the cross-crop correlation was also calculated using the third method in which the state-level corn and soybean yield data is used rather than the county-level yield data. The results for each of the states considered are as follows:

Table 10: Cross-Crop Correlation (Method 3)

State	Cross-Crop Correlation Estimate
Indiana	0.4552
Iowa	0.6438
North Carolina	0.4241

These are three of many possible methods to calculate the cross-crop correlation parameter for a state when the cross-crop correlation parameter is found not to be location, or county dependent. Which method is optimal will depend on the availability of county-level and state-level data, researcher preferences and *a priori* beliefs, among other considerations. If there are a significant number of missing county-level observations, the methods described above may provide better estimates of the cross-crop correlation when the cross-crop correlation is found not to be location dependent.

0.4 The Rate of Spatial Correlation Decay of Crop Yields

This analysis focuses on modeling the drift rate of the correlation parameter between same crop yields through space, or more specifically, the rate of spatial correlation decay of crop yields. The rate of spatial correlation decay of crop yields is the rate at which the correlation between counties' crop yields decreases as the distance between counties increases. This decay rate provides two main insights for crop insurers. First, it will provide an indication of the distance between crop yields that is required for the yields to be no longer significantly correlated. For example, if a crop insurer is insuring crop yields that are 100 miles apart, should the insurer be concerned with the potential spatial correlation of these yields? Second, it will provide an indication of degree to which the spatial correlation of crop yields has impacted the FCIC's portfolio risk and poor actuarial performance. For example, if the rate of spatial correlation is found to gradually decrease such that crop yields produced hundreds of miles apart are significantly correlated with each other, the insurer's portfolio risk is substantially greater then if the rate of spatial correlation decays at a faster rate.

With the exception of Wang[29] and Goodwin[8], little research has explored the spatial relationships of crop yields. Wang finds yield similarities among Washington State farm wheat yields and Iowa farm corn yields. The farms are then grouped into "zone-based" clusters based on their yield similarities rather than their spatial location[55]. Goodwin provides an initial exploration of the rate of spatial correlation

[55]Wang proposes these zone-based clusters as the method for grouping farms in the Group Risk Plan (GRP) rather than the currently used county-based method of grouping as a way to improve

decay by examining the rate at which the spatial correlation decays as the distance between crop yields increases. This research extends Goodwin's analysis by developing a spatial correlation model and examining its properties and characteristics. In addition, this analysis also suggests that the rate of spatial correlation decay may depend on not only on the distance that crop yields are apart, but also the direction (north/south versus east/west) that they are separated.

0.4.1 The Rate of Spatial Correlation Decay Model

The rate of spatial correlation decay model is estimated using the county-level corn yields data from Illinois, Indiana, Iowa, and North Carolina, as described in the "Data" section. For each state considered, the rate of spatial correlation decay model is developed as follows:

1. For each county x and y, the yield data for the crops m and n are detrended using the method described in the section "Method Used to Remove Time Trend".

2. For each state, the Spearman rank correlation coefficient is calculated such that $\rho_{xy,c}$ is the Spearman rank correlation coefficient between $\epsilon_{x,c}$ and $\epsilon_{y,c}$, the residual corn yield data from county x and county y, respectively.

3. The great circle distance in miles[56] between each of the counties in a given state is calculated such that d_{xy} is the great circle distance in miles between

the risk-management and cost effectiveness of the GRP.

[56]The great circle distance d_{xy} (in miles) between two points x and y is equal to: $d_{xy} = acos^{-1}[cos(\delta_x)cos(\delta_y)cos(\lambda_x - \lambda_y) + sin(\delta_x)sin(\delta_y)]3963$, where acos is the arc cosine, δ_x and δ_y are the latitude coordinates for points x and y, respectively, λ_x and λ_y are the longitude coordinates for points x and y, respectively, and 3963 is the radius of the Earth in (statute) miles. The centroid latitude and longitude coordinates of the counties were used.

county x and county y.

4. The following model is used to estimate the rate of spatial correlation decay:

$$\rho_{xy,c} = exp(\beta_1 d_{xy} + \beta_2 d_{xy}^2 + \beta_3 d_{xy}^3)\nu_{xy,c} \tag{2}$$

where $\rho_{xy,c}$ is a ((L*(L-1)/2) x 1) vector of the Spearman rank correlation coefficients between the residual corn yield data from county x and y, d_{xy} is a ((L*(L-1)/2) x 1) vector of the great circle distance in miles (scaled by 100) between county x and county y, d_{xy}^2 is a ((L*(L-1)/2) x 1) vector of the squared distance measure, d_{xy}^3 is a ((L*(L-1)/2) x 1) vector of the cubed distance measure[57], and $\nu_{xy,c}$ is a ((L*(L-1)/2) x 1) residual vector[58].

This model was selected for both theoretical and statistical reasons. In theory, we expect that the correlation between crop yields should approach zero as the distance between the yields approaches infinity. Notice that as long as the β_3 parameter estimate is negative, the Spearman rank correlation coefficient will go to zero as the distance between the corn yields goes to infinity. The model described in Eq. (2) also provided the best overall fit (compared to the other models considered) to the corn yield data of the four states considered. The criteria used to determine the best fit were the MSE, R^2, Adjusted R^2, F value, the value of the likelihood function, and the t-values of the parameter estimates. The additional models explored are listed in Appendix A.

[57] d_{xy} = .01*(great circle distance in miles between county x and county y); $d_{xy}^2 = .01^2 * d_{xy} * d_{xy}$; $d_{xy}^3 = .01^3 * d_{xy} * d_{xy} * d_{xy}$
Scaling the distance measures allows for a more clear interpretation of the parameter estimates.

[58] The model estimated that was estimated is: $ln(\rho_{xy,c}) = \beta_1 d_{xy} + \beta_2 d_{xy}^2 + \beta_3 d_{xy}^3 + ln(\nu_{xy,c})$, such that $ln(\nu_{xy,c}) = ln(\rho_{xy,c}) - \beta_1 d_{xy} - \beta_2 d_{xy}^2 - \beta_3 d_{xy}^3$

It is important to discuss how the spatial decay model in Eq. (2) should be interpreted. In this analysis, distance is used as a proxy for similarities in soil type and climate. It is likely that the similarities in soil type and climate exhibit spatial correlation which induces spatial correlation in crop yields. Distance, therefore, is used to capture the commonality in soil type and climate which is likely to cause the spatial correlation of crop yields. The model described in Eq. (2) is a method for modeling the spatial correlation of yields using the distance between crop yields to represent the spatial similarities of soil type and climate. In this light, the residual vector ν should be interpreted as representing a population variation in ρ that is not captured by distance.

The results of the estimation of Eq. (2) are as follows:

Table 11: Eq. (2) Estimation Results: Illinois

Parameter	Estimate	St. Error	t-value	Prob $> \|t\|$
$\hat{\beta}_1$	-0.5789	0.0114	-50.75	<0.0001
$\hat{\beta}_2$	0.1877	0.0113	16.59	<0.0001
$\hat{\beta}_3$	-0.0336	0.0026	-12.77	<0.0001

R^2	Adjusted R^2	MSE	F value
0.5697	0.5695	0.0313	17095.7

Table 12: Eq. (2) Estimation Results: Indiana

Parameter	Estimate	St. Error	t-value	Prob $> \|t\|$
$\hat{\beta}_1$	-0.7830	0.0211	-37.06	<0.0001
$\hat{\beta}_2$	0.3426	0.0278	12.33	<0.0001
$\hat{\beta}_3$	-0.0728	0.0086	-8.45	<0.0001

R^2	Adjusted R^2	MSE	F value
0.3893	0.3890	0.0487	8377.2

Table 13: Eq. (2) Estimation Results: Iowa

Parameter	Estimate	St. Error	t-value	Prob $> \|t\|$
$\hat{\beta}_1$	-0.5520	0.0147	-37.67	<0.0001
$\hat{\beta}_2$	0.1548	0.0171	9.03	<0.0001
$\hat{\beta}_3$	-0.0214	0.0047	-4.47	<0.0001

R^2	Adjusted R^2	MSE	F value
0.4714	0.4712	0.0302	13254.5

Table 14: Eq. (2) Estimation Results: North Carolina

Parameter	Estimate	St. Error	t-value	Prob > $\|t\|$
$\hat{\beta}_1$	-0.8292	0.0241	-34.38	<0.0001
$\hat{\beta}_2$	0.2728	0.0199	13.74	<0.0001
$\hat{\beta}_3$	-0.0412	0.0038	-10.80	<0.0001

R^2	Adjusted R^2	MSE	F value
0.2510	0.2507	0.2240	4706.1

Graphs of each state's estimated model overlaying the data points are shown in the following figures. A significance line showing the values at which the Spearman rank correlation coefficient is significantly different than zero, is provided on each graph[59]. Notice that the distance scale varies by state.

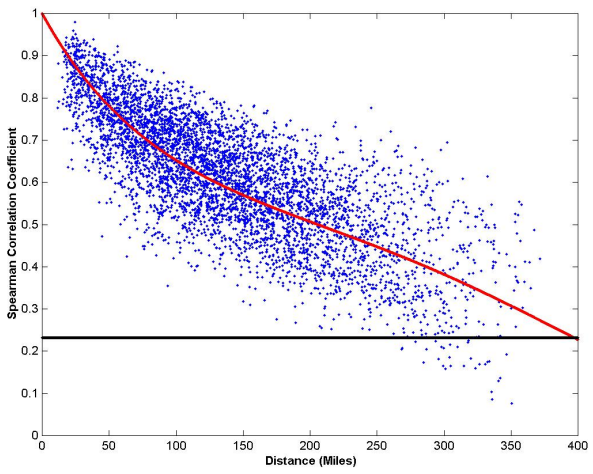


Figure 22: Rate of Spatial Correlation Decay Model: Illinois

[59]The critical value for the null hypothesis $H_o : \rho_{xy,c} = 0$ is 0.2326 for Indiana, Illinois, and Iowa and 0.2479 for North Carolina such that any correlation value greater than its respective critical value is considered significantly different from zero.

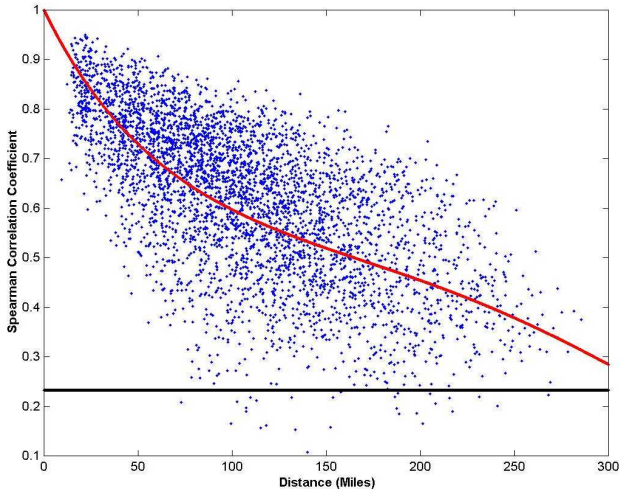


Figure 23: Rate of Spatial Correlation Decay Model: Indiana

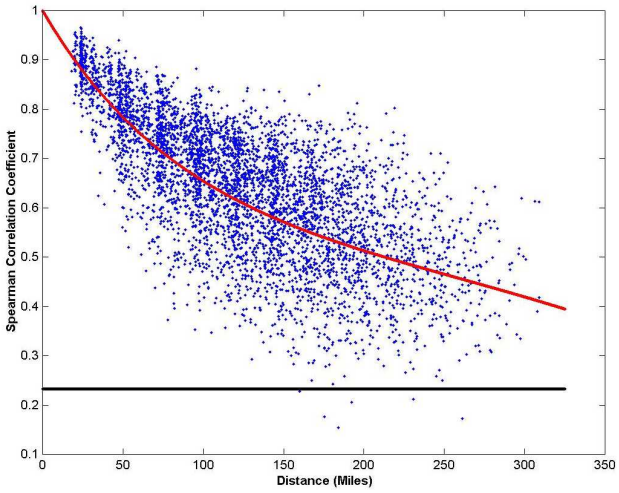


Figure 24: Rate of Spatial Correlation Decay Model: Iowa

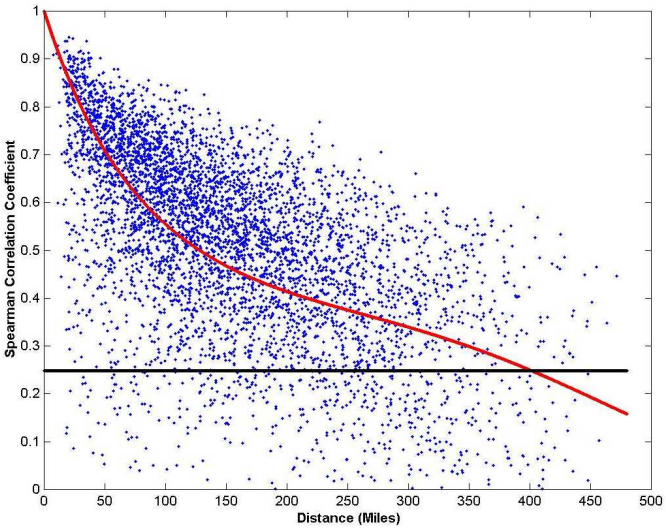


Figure 25: Rate of Spatial Correlation Decay Model: North Carolina

For all states considered, the $\hat{\beta}_3$ parameter estimate is negative. This result ensures that the Spearman rank correlation coefficient will go to zero as the distance between the yields goes to infinity. Additionally, each of the following null hypothesis were tested using the LM test[60]:

- $\mathrm{H}_o = \beta_1 = \beta_2 = \beta_3 = 0$
- $\mathrm{H}_o = \beta_1 = \beta_2 = 0$
- $\mathrm{H}_o = \beta_1 = \beta_3 = 0$
- $\mathrm{H}_o = \beta_2 = \beta_3 = 0$
- $\mathrm{H}_o = \beta_1 = 0$
- $\mathrm{H}_o = \beta_2 = 0$

[60]The null hypothesis were also tested using the LR and Wald tests, which provided the same results are the LM test.

- $H_o = \beta_3 = 0$

For each state, the LM test rejected each of the above null hypothesis.

Positive Definiteness of the Variance/Covariance Matrix

A concern with the model described in Eq. (2) is the positive definiteness of the variance/covariance matrix in distance. A proof of the positive definiteness of the variance/covariance matrix is beyond the scope of this paper. However, a simulation study was performed to test for positive definiteness over a large geographical area that encompasses the states considered in this analysis. The simulation study is described as follows:

1. 100 pairs of values were randomly drawn from the normal distributions N(0.6,0.2) and N(1.5,0.2). These pairs represent the latitude and longitude coordinates, respectively, in radians[61].

2. The great circle distance was computed for each pair of randomly drawn latitude and longitude coordinates which generated the d_{xy} elements, creating the (100 x 100) distance matrix D. The squared and cubed distance measures d_{xy}^2 and d_{xy}^3 were computed and these distance measures became the elements of the (100 x 100) matrices D^2 and D^3, respectively.

3. For each state, the parameter estimates were used to estimate the Ω matrix such that[62]:

$$\Omega = exp(\beta_1 D + \beta_2 D^2 + \beta_3 D^3)$$

[61]The means and variances were chosen such that they would cover a large geographical area (approximately one-third of the United States) and encompass the states considered in this analysis.

[62]The exponential function refers to taking the exponential value element-wise and not the exponential of the matrix.

4. The sign definiteness of the Ω matrix was determined by the sign of the minimum eigenvalue of Ω. As long as the minimum or smallest eigenvalue is greater than zero, Ω is considered to be positive definite.

5. The above steps were repeated 50,000 times for each state considered.

For each of the states considered and the parameter estimates reported previously, the Ω matrix was found to always be positive definite and is positive definite using the parameter estimates and the actual distance matrices. Although not a proof, these results provide an indication that the variance/covariance matrix in distance is positive definite.

Properties of the Rate of Spatial Correlation Decay Model

The graphs of the spatial decay model provide a visual picture of how the correlation between two crop yields decreases as the distance between the yields increases. However, given the large number of observations on each graph, it is difficult to determine the average correlation value at a particular distance. The following table provides this information by giving the average Spearman rank correlation coefficient over various distance ranges (in miles).

Table 15: Average Spearman Correlation Over Distance Ranges

	Illinois	Indiana	Iowa	North Carolina
Distance	$\bar{\rho}$	$\bar{\rho}$	$\bar{\rho}$	$\bar{\rho}$
0-50	0.8138	0.7830	0.8439	0.7129
51-100	0.7047	0.6693	0.7131	0.6220
101-150	0.6263	0.5787	0.6247	0.5427
151-200	0.5520	0.5072	0.5590	0.4891
201-250	0.4899	0.4517	0.5067	0.4337
251-300	0.4258	0.3704	0.4708	0.3712
301-350	0.3770	.	.	0.3341
351-400	.	.	.	0.3248
401-451	.	.	.	0.2670

The above table indicates that as the distance increases, the Spearman rank correlation coefficient decreases, as is expected. However, these measures depend on monotoncity for them to be unique. In order for the rate of spatial correlation decay model to exhibit strict decreasing monotoncity, the first derivative of the model with respect to distance[63] must be less than zero such that the following inequality holds for all distances:

$$\beta_1 + 2\beta_2 d_{xy} + 3\beta_3 d_{xy}^2 < 0$$

This inequality yields the following necessary conditions for the model parameters:

- β_1, β_2, and $\beta_3 < 0$

  or

- $\beta_1, \beta_3 < 0$ and $3\beta_1\beta_3 > \beta_2^2$, for $\beta_2 > 0$

The above conditions were evaluated using each state's rate of spatial correlation

[63] $$\frac{\partial \rho_{xy,c}}{\partial d_{xy}} = (\beta_1 + 2\beta_2 d_{xy} + 3\beta_3 d_{xy}^2)\rho_{xy,c}$$

decay model parameter estimates. For each of the state's model parameters, the above conditions held, indicating that the rate of spatial correlation decay model is monotonically decreasing.

Rate of Spatial Correlation Decay Dependence on Distance and Direction

The model described in Eq. (2) assumes that the rate of spatial correlation decay is independent of the distance direction (north/south versus east/west). In other words, crop yields from two counties that are 50 miles apart in a north/south direction are treated the same as crop yields from two counties that are 50 miles apart in a east/west direction. However, it is likely that the rate of spatial correlation decay depends on the direction as well as the distance that the crop yields are separated. For example, Iowa is more homogeneous in terms of weather and soil patterns in an east/west direction than a north/south direction. These weather and soil patterns could lead to a stronger crop yield correlation in an east/west direction compared to a north/south direction. To illustrate this concept, the following figures contain color maps that select a single county and show how the corn yield correlation changes between the selected county and all other counties in that state[64].

[64]Two counties are selected randomly from each state, creating eight color maps. In each map, the county that is colored in the darkest shade of red is the selected county.

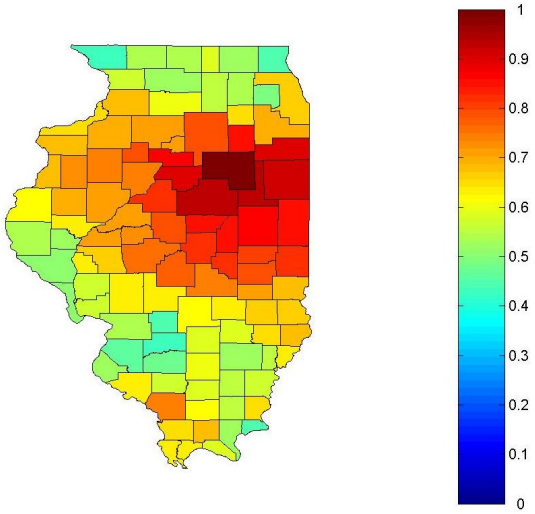


Figure 26: Illinois: Corn Yield Correlations: Livingston County

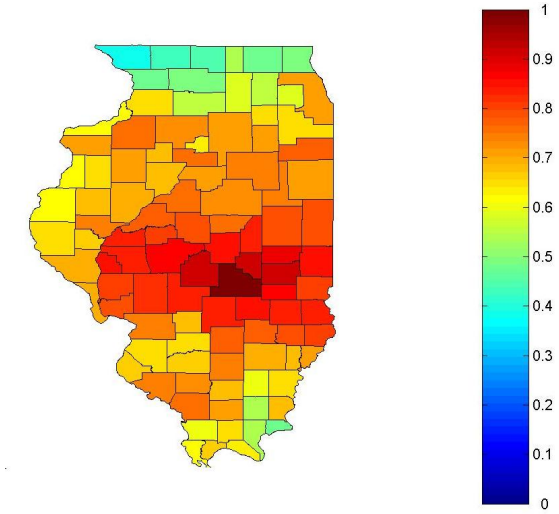


Figure 27: Illinois: Corn Yield Correlations: Shelby County

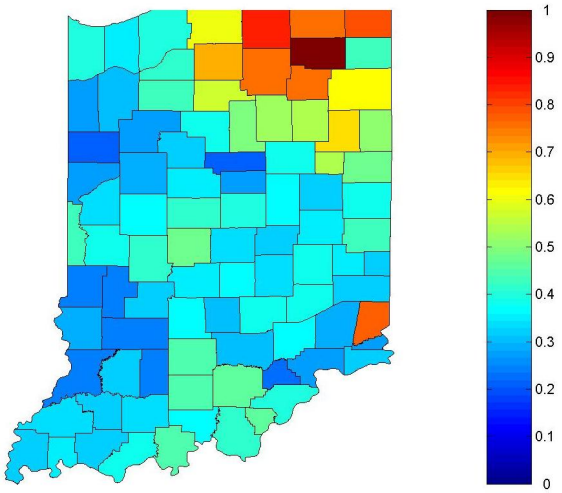


Figure 28: Indiana: Corn Yield Correlations: Noble County

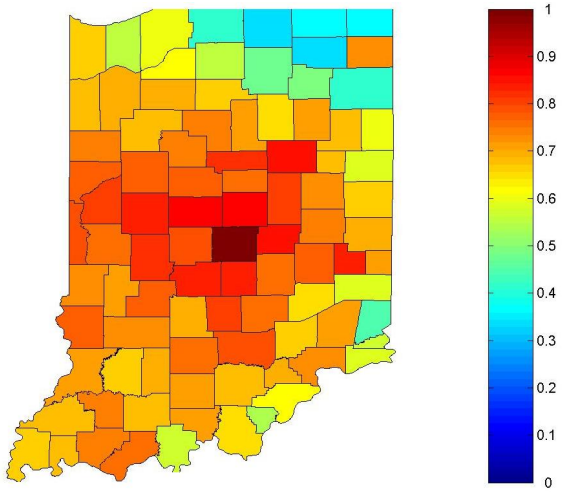


Figure 29: Indiana: Corn Yield Correlations: Marion County

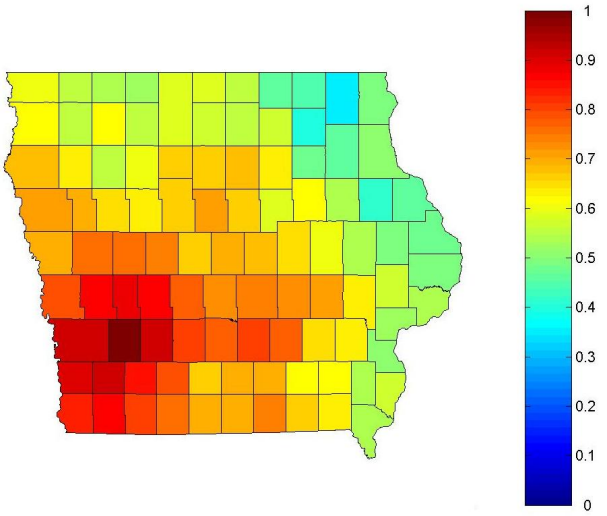


Figure 30: Iowa: Corn Yield Correlations: Cass County

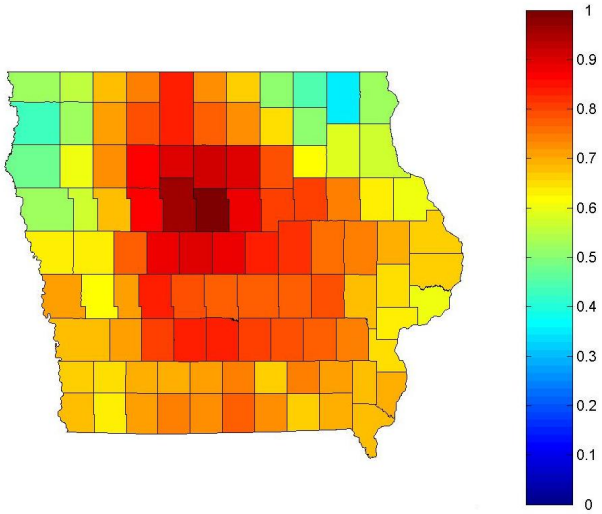


Figure 31: Iowa: Corn Yield Correlations: Hamilton County

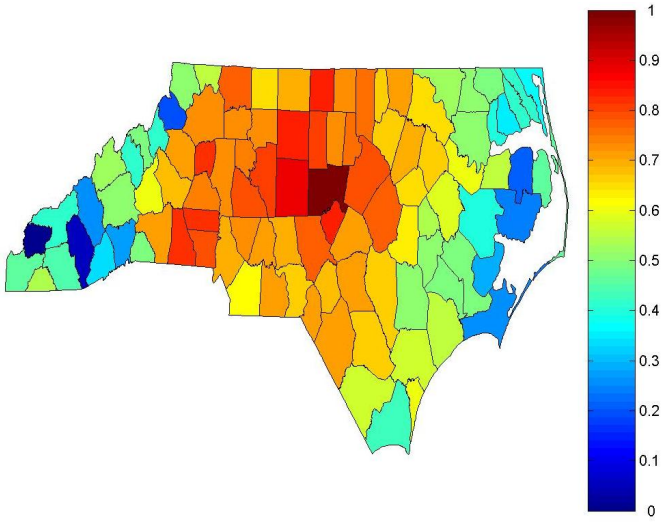


Figure 32: North Carolina: Corn Yield Correlations: Chatham County

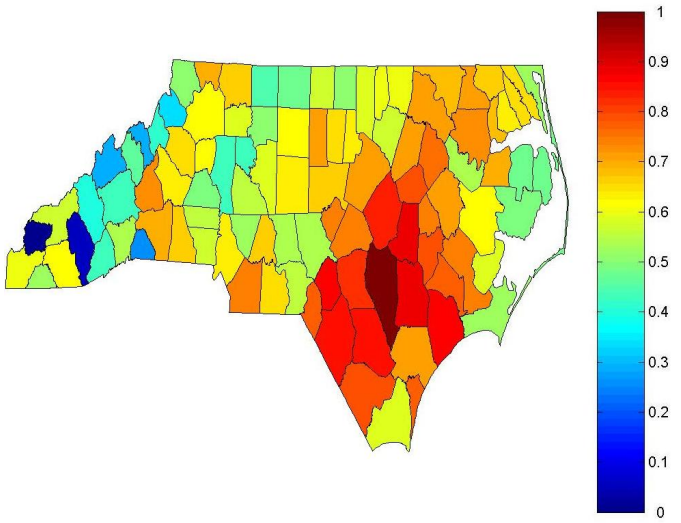


Figure 33: North Carolina: Corn Yield Correlations: Sampson County

Some of the graphs indicate that the corn yield correlation tends to be stronger in a north/south direction compared to an east/west direction such that the correlation tends to decay more gradually in a north/south direction as compared to an east/west direction. An example of this is illustrated in Figure 33. Some of the graphs indicate that the corn yield correlation tends to be stronger in an east/west direction compared to a north/south direction such that the correlation tends to decay more gradually in an east/west direction as compared to a north/south direction. An example of this is illustrated in Figure 27. In other graphs, the correlation between corn yields appears to decay at an approximately the same rate. An example of this is illustrated in Figure 32.

To further explore this issue, the following model extends Eq. (2) by allowing for the rate of spatial correlation decay to depend on direction as well as the distance that crop yields are separated[65].

$$\rho_{xy,c} = exp(\beta_1 d_{xy} + \beta_2 d_{xy}^2 + \beta_3 d_{xy}^3 + \beta_4 \omega_{xy} + \beta_5 \omega_{xy}^2 + \beta_6 \omega_{xy}^3)\nu_{xy,c} \tag{3}$$

where ω_{xy} is a ((L*(L-1)/2) x 1) vector of the north/south distance that the counties x and y are apart in miles scaled by 100, and remainder of the variables and parameters are the same as described previously[66]. The ω_{xy} variables represent the degree to which counties x and y are separated in a north/south direction such that the ω_{xy} variables can be interpreted as a "north/south distance" measure between the counties. Therefore, Eq. (3) allows for the rate of the spatial correlation decay

[65]Recall that Eq. (2) provided the best overall fit to the county-level corn yield data from the four states considered in this analysis. Eq. (3) is chosen because it nests Eq. (2).

[66]ω_{xy} = .01*(north/south distance between county x and county y in miles); $\omega_{xy}^2 = .01^2 * \omega_{xy} * \omega_{xy}$; $\omega_{xy}^3 = .01^3 * \omega_{xy} * \omega_{xy} * \omega_{xy}$

of crop yields to depend on the "distance direction" as well as the total distance between crop yields.

Eq. (3) was estimated using the county-level corn yield data from the four states considered in this analysis. The results of the estimation of Eq. (3) are as follows:

Table 16: Eq. (3) Estimation Results: Illinois

Parameter	Estimate	St. Error	t-value	Prob $> \|t\|$
$\hat{\beta}_1$	-0.5781	0.0190	-30.50	<0.0001
$\hat{\beta}_2$	0.1753	0.0145	12.07	<0.0001
$\hat{\beta}_3$	-0.0299	0.0031	-9.65	<0.0001
$\hat{\beta}_4$	-0.0308	0.0412	-0.75	0.4531
$\hat{\beta}_5$	-0.0704	0.0617	-1.14	0.2540
$\hat{\beta}_6$	-0.0141	0.0262	-0.54	0.5902

R^2	Adjusted R^2	MSE	F value	Likelihood Value
0.5737	0.5733	0.0310	8632.4	159.7489

Table 17: Eq. (3) Estimation Results: Indiana

Parameter	Estimate	St. Error	t-value	Prob $> \|t\|$
$\hat{\beta}_1$	-0.8300	0.0354	-23.48	<0.0001
$\hat{\beta}_2$	0.3698	0.0358	10.34	<0.0001
$\hat{\beta}_3$	-0.0781	0.0101	-7.67	<0.0001
$\hat{\beta}_4$	0.0518	0.0694	0.74	0.4558
$\hat{\beta}_5$	0.0098	0.1235	0.08	0.9356
$\hat{\beta}_6$	-0.0203	0.0621	-0.33	0.7429

R^2	Adjusted R^2	MSE	F value	Likelihood Value
0.3911	0.3904	0.0486	4199.65	203.0704

Table 18: Eq. (3) Estimation Results: Iowa

Parameter	Estimate	St. Error	t-value	Prob > $\|t\|$
$\hat{\beta}_1$	-0.7027	0.0276	-25.39	<0.0001
$\hat{\beta}_2$	0.2790	0.0272	10.26	<0.0001
$\hat{\beta}_3$	-0.0672	0.0077	-8.90	<0.0001
$\hat{\beta}_4$	0.2037	0.0336	6.07	<0.0001
$\hat{\beta}_5$	-0.1658	0.0341	-4.87	<0.0001
$\hat{\beta}_6$	0.0609	0.0097	6.23	<0.0001

R^2	Adjusted R^2	MSE	F value	Likelihood Value
0.4973	0.4967	0.0287	7004.92	139.0836

Table 19: Eq. (3) Estimation Results: North Carolina

Parameter	Estimate	St. Error	t-value	Prob > $\|t\|$
$\hat{\beta}_1$	-1.7045	0.0945	-17.33	<0.0001
$\hat{\beta}_2$	0.4892	0.0922	5.30	<0.0001
$\hat{\beta}_3$	-0.0754	0.0243	-3.10	0.0019
$\hat{\beta}_4$	0.2444	0.0976	2.50	0.0125
$\hat{\beta}_5$	-0.2286	0.0918	-2.49	0.0128
$\hat{\beta}_6$	0.0373	0.0244	1.53	0.1250

R^2	Adjusted R^2	MSE	F value	Likelihood Value
0.2529	0.2522	0.2236	2359.69	1105.0

Eq. (3) provided a good fit to the Iowa county-level corn yield data, indicating that the rate of spatial correlation decay is dependent on the direction as well as the distance that crop yields are separated in Iowa. However, the north/south "direction parameters" ($\hat{\beta}_4$, $\hat{\beta}_5$, and $\hat{\beta}_6$) are not significant for the Indiana and Illinois data. Only the $\hat{\beta}_6$ parameter is not significant for the North Carolina data.

Model Discussion and Interpretation of Estimation Results

Perhaps the best method of interpreting the rate of spatial correlation models with dependence on distance and direction is through graphical representation. The following are 2-dimensional graphs of the Spearman rank correlation coefficient in both total distance (in miles) d_{xy} and the north/south distance (in miles) ω_{xy} using Iowa and North Carolina estimation results[67].

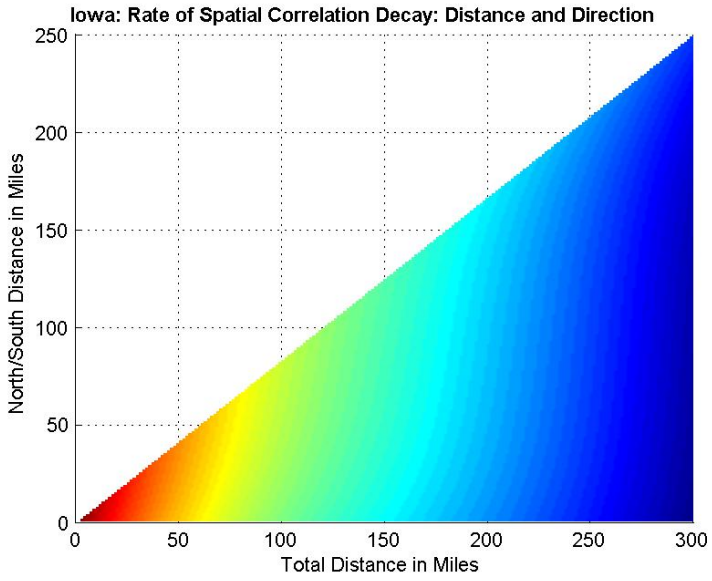


Figure 34: Spatial Correlation Decay: Distance and Direction (Iowa)

[67]Note that Eq. (3) is undefined above the 45 degree line as north/south distance cannot exceed total distance, and, as a result, these points are not plotted.

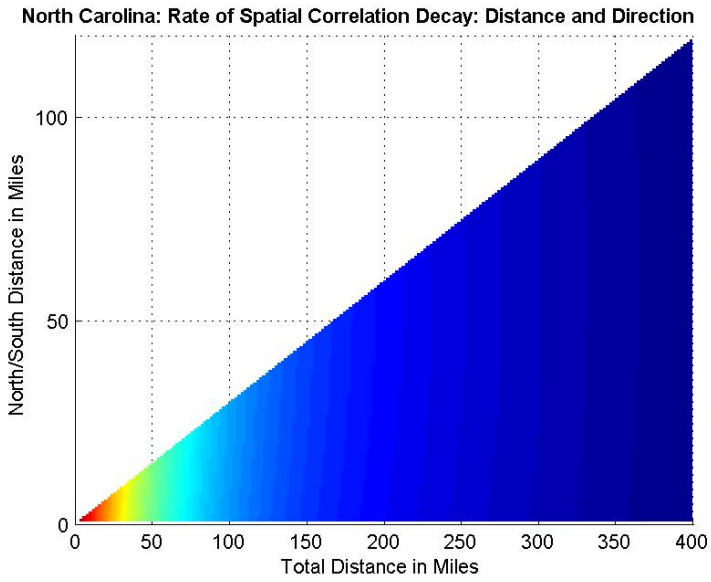


Figure 35: Spatial Correlation Decay: Distance and Direction (North Carolina)

The Iowa graph shows that the greatest correlation between corn yields is found when both the total distance and north/south distance are minimized. The correlation between corn yields tends to gradually decrease as both total distance and north/south increase, but the correlation tends to decrease at a faster rate as total distance increases compared to the rate at which it decreases as north/south distance increases. This graph indicates that the correlation is the greatest when corn yields are separated by less than 50 miles in both total distance and north/south distance.

Similar to the Iowa graph, the North Carolina graph shows that the greatest correlation between corn yields is found when both the total distance and north/south distance are minimized. However, the correlation between the corn yields in North Carolina seems to decay more rapidly than the correlation between the Iowa corn yields. This result coincides with the fact that Iowa is more homogenous in terms of weather and soil patterns than North Carolina, which is likely to lead to a slower

correlation decay between the corn yields in Iowa compared to North Carolina. Further, the correlation appears to decay at approximately the same rate as both total distance and north/south distance increase. This graph indicates that the correlation is the greatest when corn yields are separated by less than 25 miles in both total distance and north/south distance.

Although these graphs provide a good visual interpretation of the estimation results, one cannot determine from the graphs if the correlation decays to zero as the north/south distance measure increases, or if the correlations are strictly monotonically decreasing in ω_{xy}. A discussion of each follows.

In theory, we expect that the correlation between crop yields should approach zero as the distance measure between yields approaches infinity. In order to obtain this result, the following parameter conditions must hold for Eq. (3):

$$|\beta_3| > |\beta_6| \text{ where } \beta_3 < 0$$

The above parameter conditions hold for both the Iowa and North Carolina parameter estimates. This indicates that the correlation between corn yields approaches zero as the distance between yields approaches infinity in Iowa and North Carolina. These results correspond to the graphical representation of Eq. (3) shown in Figure 34 and Figure 35, respectively.

A second issue to consider in the analysis of these graphs is whether or not the correlation is strictly monotonically decreasing in distance. In order for Eq. (3) to be strictly monotonically decreasing, the first derivative of the model with respect

to distance[68] must be less than zero such that the following holds for all d_{xy} and ω_{xy}:

$$\beta_1 + 2\beta_2 d_{xy} + 3\beta_3 d_{xy}^2 + \beta_4 + 2\beta_5 \omega_{xy} + 3\beta_6 \omega_{xy}^2 < 0$$

If α_1, α_2, and α_3 are defined as:

$$\alpha_1 = \beta_1 + \beta_4 \qquad \alpha_2 = 2(\beta_2 + \beta_5) \qquad \alpha_3 = 3(\beta_3 + \beta_6)$$

The inequality for Eq. (3) yields the following necessary conditions for the model parameters:

$$|\alpha_3| < |(\alpha_1 + \alpha_2)| \text{ where } \alpha_3 < 0$$

The necessary conditions for monotoncity were checked using the Iowa and North Carolina parameter estimates for Eq. (3). The conditions held for both the Iowa and North Carolina parameter estimates, indicating that Eq. (3) is strictly monotonically decreasing in distance in both states. This result corresponds to the graphical representation of Eq. (3) shown in Figure 34 and Figure 35, respectively.

A final issue to consider is whether or not the variance/covariance matrices defined by the Iowa and North model estimates are positive definite in both distance and the north/south distance measure. A simulation was performed similar to the one described in the section "Positive Definiteness of the Variance/Covariance Matrix" in this paper with the exception that the parameter estimates of Eq. (3) for Iowa and North Carolina were used. The Ω matrix was computed using the Iowa parameter

[68] $$\frac{\partial \rho_{xy,c}}{\partial \omega_{xy}} = (\beta_1 + 2\beta_2 d_{xy} + 3\beta_3 d^2 + \beta_4 + 2\beta_5 \omega_{xy} + 3\beta_6 \omega_{xy}^2)\rho_{xy,c}$$

estimates and then the North Carolina parameter estimates. Eq. (3) was found not to be positive definite in at least one of the replications using the Iowa parameter estimates. However, in each of the replications, Eq. (3) was positive definite using the North Carolina parameter estimates and, although not a proof, these results suggest that variance/covariance matrix of Eq. (3) is positive definite in distance using the North Carolina model estimates.

0.4.2 Conclusion of the Rate of Spatial Correlation Analysis

Several rate of spatial correlation decay models were developed and tested. Eq. (2) provided the best overall fit to the county-level corn yield data from the four states considered in this analysis. Further, each of the state's parameter estimates indicated that the Spearman rank correlation coefficient will go to zero as the distance between the corn yields goes to infinity. The sign definiteness of the variance/covariance matrix in distance of Eq. (2) was explored. Although a proof of the sign definiteness of this matrix is beyond the scope of this paper, the simulation study performed indicated that the variance/covariance matrix is likely to be positive definite in distance. The property of monotoncity was examined. For each of the states considered, the estimated rate of spatial correlation decay model is strictly monotonically decreasing in distance.

Finally, Eq. (2) was extended to allow for the rate of spatial correlation decay model to depend on the direction as well as the distance that the crop yields are separated. For the Iowa and North Carolina corn yield data, the dependence of the rate of spatial correlation decay model on north/south distance as well as the total distance was significant. The parameter estimates for the Iowa and North Carolina models

indicated that Eq. (3) is strictly monotonically decreasing in distance. Further, the simulation performed to test the positive definiteness of the variance/covariance matrix of Eq. (3) indicated that the variance/covariance matrix for the Iowa model was not positive definite in distance, but the simulation results suggested that the North Carolina model is likely to positive definite in distance.

Many implications for crop insurers can be drawn from this rate of spatial correlation decay analysis. Consider the rate of spatial correlation graphs presented previously. The county-level crop yields from the states considered in this analysis are significantly correlated to distances as great as 475 miles[69]. Therefore, the size of the geographical area that crop insurers need to consider that yields may be spatially correlated is at least a 475 mile radius. Further, because the geographical area that contains correlated crop yields is relatively large, this indicates that the spatial correlation of crop yields is likely to be a significant contributor to the FCIC's poor actuarial performance.

[69]The critical value for the null hypothesis $H_o : \rho_{xy,c} = 0$ is 0.2326 for Indiana, Illinois, and Iowa and 0.2479 for North Carolina. The majority of the Spearman rank correlation coefficients are greater than their appropriate critical values at the greatest distance measures possible for each state.

0.5 The Bivariate Distributions of Crop Yields: The Copula Methodology

This final analysis section explores the bivariate distributions of crop yields. These bivariate distributions allow for an analysis of the higher-order spatial relationships, or higher-order cross-moments of crop yields, and they can be used to develop crop insurance premiums that account for the spatial correlation of crop yields. The purpose of this analysis is to determine the "best" bivariate distribution for crop yields and to determine if it is appropriate to use one bivariate distribution over a geographical area, in particular, the Normal bivariate copula. In a more general context, this analysis seeks to determine if the parameters of the bivariate distribution of crop yields change slowly enough over a particular region that the bivariate distribution of crop yields can be considered the same over that area.

Perhaps the best method for estimating these bivariate distributions is to use the concept of a copula. A copula is a distribution on the unit square in which its marginal distributions are uniform. Haas[11] states that copulas provide a method for describing the correlation structure between random variables that is independent of their individual marginals. This implies that the joint behavior of two or more correlated random variables with any given individual marginal distributions can be represented using the copula methodology and the correlation between the random variables and the individual marginal distributions are maintained in the joint representation. Copulas are defined as follows:

$$H(u) = C(F_1(u_1), F_2(u_2), ...F_n(u_n))$$

where H(.) is a multivariate distribution function comprised of the uniform marginal distributions $F_1(u_1), F_2(u_2), ...F_n(u_n)$, and C(.) is defined as the copula. By Sklar's Theorem, for any distribution function H(.) with uniform marginals $F_1(u_1), F_2(u_2), ...F_n(u_n)$ there *always exists* a copula C(.) such that the above relationship holds. Sklar[25] also proved the converse such that if C(.) is a copula, then H(.) *is always* a distribution function such that the above relationship holds.

Bivariate copulas are used in this analysis because, with the exception of the multivariate Normal copula and *t*-copula, multivariate copulas are unknown or quickly become untractable. The bivariate copula is defined as follows:

$$H(m, n) = C(F_m(m), F_n(n))$$

where H(.) is a bivariate distribution comprised of two uniform marginal distributions, and C(.) is the bivariate copula.

All that is required to fit a bivariate set of random variables to a bivariate copula is to transform the two sets of random variables to their individual CDF values. The CDF values can then used to fit the bivariate copula using maximum likelihood estimation. In other words, by using the copula methodology, any two correlated random variables with any individual marginal distributions[70] can be represented as a joint distribution while maintaining the correlation and the individual marginal distributions. However, there are many different copula representations available, such that two random variables with a specified Spearman rank correlation coefficient and individual marginal distributions can be represented differently given the functional form of the copula. In recent literature, Haas[11], Frees and Valdez[2] have

[70]The individual marginal distributions need to be continuous.

demonstrated that the choice of the functional form of the copula has a significant impact on the shape of the distribution. These differences are particularly noticeable in the corners of the joint distributions. The corners, especially the lower corner, of crop yield distributions are important to crop insurers.

As a demonstration of the differences in the functional form of bivariate copulas, density plots of several copulas follow. The same Spearman rank correlation coefficient (0.6) between the random variables m and n is represented in each of the density plots. The bivariate copula distributions used in this analysis are members of common bivariate copula families discussed in the literature. Appendix B (Part One) contains a functional form definition of all the bivariate copulas used in this analysis.

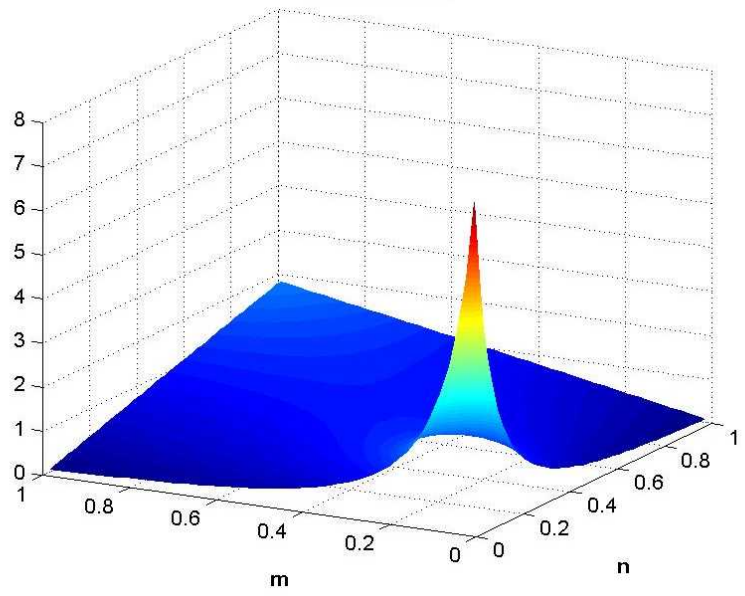


Figure 36: Ali-Mikhail-Haq Copula

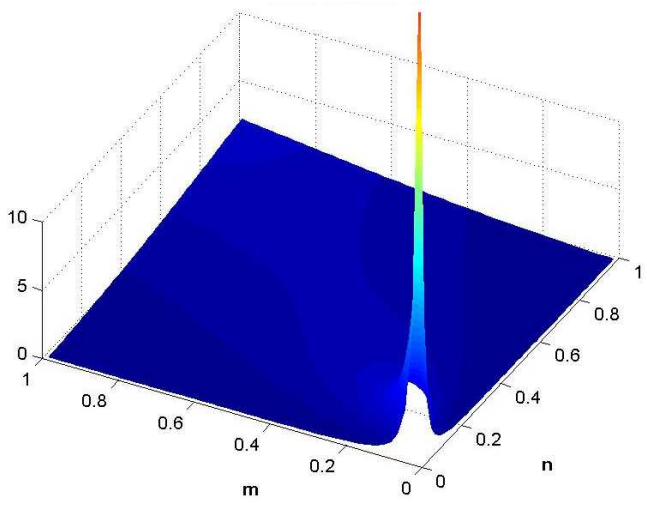


Figure 37: Cook-Johnson Copula

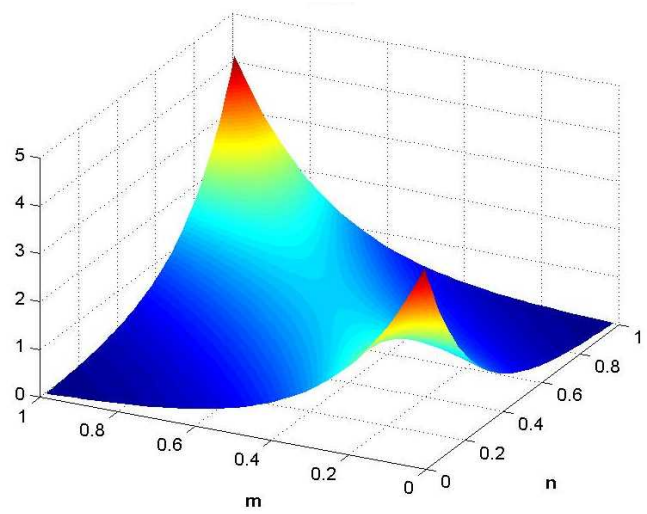


Figure 38: Frank Copula

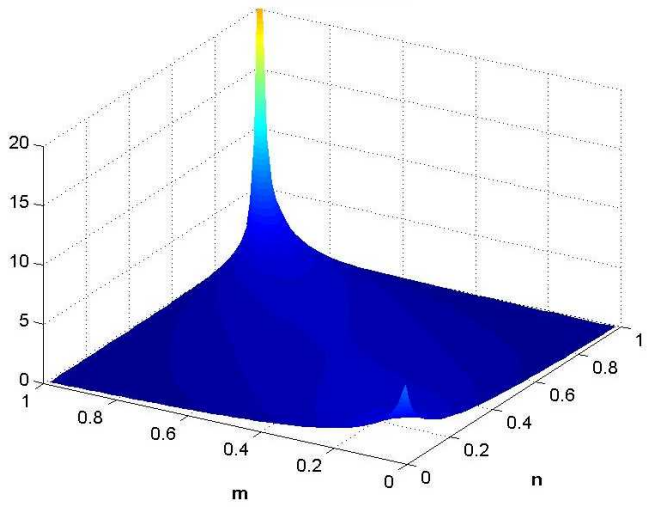


Figure 39: Gumbel-Hougaard Copula

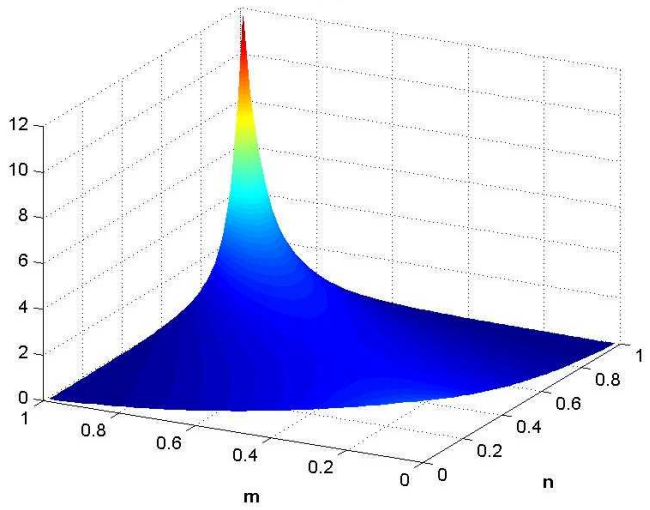


Figure 40: Joe Copula

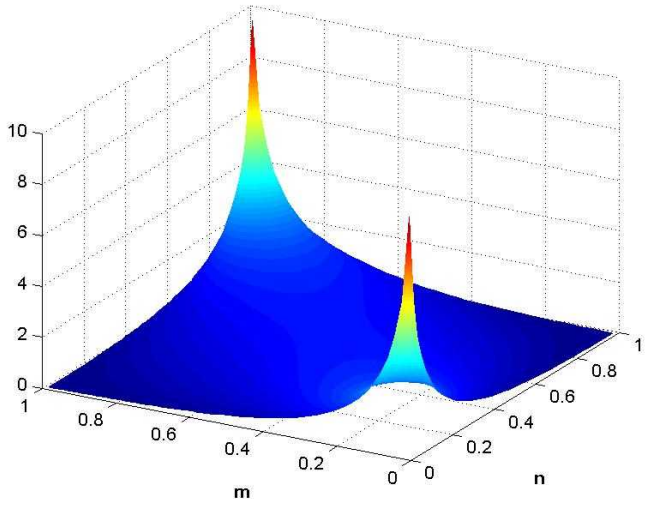


Figure 41: Normal Copula

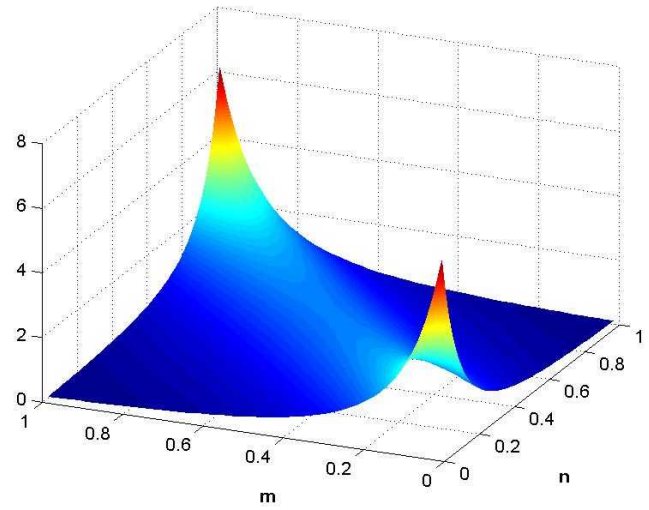


Figure 42: Plackett Copula

Again, each of these copula density plots were created using the same two random variables with a Spearman rank correlation coefficient of 0.6. Each copula represents the joint behavior of the random variables while maintaining the individual marginal distributions and the correlation between the random variables. However, each joint representation is different, depending on the functional form of the copula.

The bivariate copula density plots clearly illustrate that the choice of the functional form of the copula has a dramatic impact on the shape of the joint distribution, especially in the corners of the distributions. For example, notice the difference in the areas under the densities near the extreme corner points (0,0) and (1,1). The area under the densities at the low corner (0,0) represents the probability of an extreme low/low event occurring, that is, the probability that both random variables take on an extreme low value. The area under the densities at the extreme high corner (1,1) represents the probability of an extreme high/high event occurring, that is, the probability that both random variables take on an extreme high value. The Ali-Mikhail-Haq and Cook-Johnson copulas place a greater probability on extreme low/low events occurring compared to the other copulas considered in this analysis. The Gumbel-Hougaard and Joe copulas place a greater probability on extreme high/high events occurring. The remaining three copulas, Frank, Normal, and Plankett, behave similarly in that they all place an approximately equal probability on extreme low/low events and extreme high/high events because the area underneath each of the respective densities at the extreme low corner is approximately equal to the area underneath the density at the extreme high corner. Of these three, the Normal copula places the greatest probability on extreme low/low and extreme

high/high events.

It should be noted that while it is possible to compare the probabilities of extreme events[71] occurring across the copulas, it is considerably more difficult to make such comparisons of the probabilities of other events occurring by simply looking at the density plots. For example, it would be difficult to compare the probability that both random variables m and n take on values less than or equal to 0.3 across the copulas. This issue is explored further in the next section "Copula Methodology: Crop Insurance Application" in which comparisons of probabilities are made by evaluating the copula CDFs.

A second distinguishing feature of these bivariate copula density plots is the area under the density towards the corners (0,1) and (1,0). The area under the densities in these corners can be interpreted as the probability of low/high and high/low events occurring, respectively. If there is a higher probability that extreme low/low and high/high events occur, this implies that the area under the density near the (0,1) and (1,0) corners should be relatively flat[72]. While the height of the densities at particular points varies depending on the copula, the height of the densities is the greatest at the (0,0) and (1,1) corners of the distributions and away from the (0,1) and (1,0) corners. This result implies that there is a low probability of extreme low/high or extreme high/low events occurring at this correlation value (0.6). The density plots clearly indicate that the choice of the functional form of the copula has a significant impact on shape of the joint distribution, especially in the (0,0) and

[71] An extreme event is defined as an event in which both random variables either take on extreme low values or extreme high values.

[72] Recall that the correlation between the random variables is held constant in these density plots.

(1,1) corners. Given that the shape of the joint distribution, especially the corners, is of primary interest, it is important to select the copula that "best" fits the random variables.

The following procedure can be used to find the bivariate copula that "best" fits a bivariate set of random variables. First, the individual CDF values of each bivariate set of random variables is estimated. Next, the bivariate set of CDF values are used to fit the copula density by maximum likelihood estimation, such that:

$$ln\pounds = \Sigma_{i=1}^{n} ln[c(u_i, v_i \mid \beta)]$$

where

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}$$

β is/are the unknown copula parameter(s) to be estimated by finding the β that maximizes $\pounds$, and u and v represent the CDF values of the random variables. This process can be repeated for each bivariate copula considered. The copula with the greatest log likelihood value is considered the copula that best fits the random variables.

0.5.1 Application to Corn Yields

The copula methodology was applied to the county-level corn yield data, as described in the "Data" section, for each of the four states considered in this analysis.

The following describes the procedure used:

1. For each county l, the corn yield data c are detrended using the method described in the section "Method Used to Remove Time Trend" and the $\epsilon_{l,c}$ residuals are retained.

2. The kernel transform method with fixed bandwidth is used to estimate the CDFs of the crop yields, denoted $\hat{F}_{l,c}(\epsilon_{l,c})$.

3. The CDF estimates $\hat{F}_{l,c}(\epsilon_{l,c})$ of two counties' corn yields are used to estimate each of the bivariate copulas using maximum likelihood estimation such that:

$$ln\pounds_f = \Sigma_{i=1}^{n} ln[c(\hat{F}_{x,c}(\epsilon_{x,c}), \hat{F}_{y,c}(\epsilon_{y,c}) \mid \hat{\beta})]$$

where $\pounds_f$ represents the likelihood function for the bivariate copula density f, $\hat{F}_{x,c}(\epsilon_{x,c})$ are the CDF estimates of the corn yields from county x, $\hat{F}_{y,c}(\epsilon_{y,c})$ are the CDF estimates of the corn yields from county y, and $\hat{\beta}$ is/are the unknown copula parameters to be estimated by finding the $\hat{\beta}$ that maximizes $\pounds_f$.

4. The above steps were repeated for each unique pair of county-level corn yields in each state for each bivariate copula considered in this analysis. For each unique pair of county-level corn yields, the bivariate copula with the greatest log likelihood function value was retained.

Corn Yield Results

The results of the bivariate copula methodology application to the county-level corn yields are presented below. The first set of tables present the overall results, that is, for each state, the number of times a bivariate copula was selected as the copula that best fit the corn yield data is listed in absolute and percentage terms. The second set of tables divide the results by distance, that is, for each state, the number of times a bivariate copula was selected as the copula that best fit the corn yield data from

two counties that are a particular distance apart is listed in absolute and percentage terms.

Table 20: Copula Results: Corn Yield (Absolute)

Bivariate Copula	Illinois	Indiana	Iowa	North Carolina
Ali-Mikhail-Haq	173	126	33	266
Cook-Johnson	1082	1387	265	894
Frank	702	565	310	623
Gumbel-Hougaard	342	129	1259	616
Joe	457	239	313	842
Normal	1925	1319	1970	1231
Plackett	470	421	701	478

Table 21: Copula Results: Corn Yields (Percentage)

Bivariate Copula	Illinois	Indiana	Iowa	North Carolina
Ali-Mikhail-Haq	3.4	3.0	0.7	5.4
Cook-Johnson	21.0	33.1	5.5	18.1
Frank Copula	13.6	13.5	6.4	12.6
Gumbel-Hougaard	6.6	3.1	26.0	12.4
Joe	8.9	5.7	6.5	17.0
Normal	37.4	31.5	40.6	24.9
Plackett	9.1	10.1	14.5	9.7

The two bivariate copulas that are most often selected as the best fitting copula are the Normal and Cook-Johnson copulas. The Gumbel-Hougaard copula is frequently selected as the best copula using the Iowa yield data, but was not selected as frequently in the other states. From the bivariate copula density plots presented earlier, Normal, Cook-Johnson, and Gumbel-Hougaard copulas have some distinctive features describing the joint relationship of the corn yields.

The Normal copula places an approximately equal probability on extreme low/low events and extreme high/high events such that the area under the density near the point (0,0) is approximately equal to the area under the density near the point (1,1).

In regard to the corn yield data, this means that if an extreme high corn yield is observed in one county, there is a considerably greater probability that an extreme high corn yield will also be observed in another county within that state, compared to the probability that a low yield will be observed. And, if an extreme low corn yield is observed in one county, there is a considerably greater probability that an extreme low corn yield will also be observed in another county in that state, compared to the probability that a high yield will be observed. Further, notice that the Normal copula plot indicates that the area under the density becomes almost flat as it moves towards the (0,1) and (1,1) corners. This implies that there is a very low probability of a high yield/low yield or low yield/high yield event occurring compared to the probability of a high yield/high yield or low yield/low yield event occurring.

The Cook-Johnson copula places a greater probability on extreme low/low events occurring. In regard to the corn yield data, this means that if an extreme low corn yield is observed in one county, there is a considerably greater probability that an extreme low corn yield will also be observed in another county within that state, compared to the probability that a high yield will be observed. The Cook-Johnson density plot shows that the height of the density increases towards the high corner (1,1), although the greatest height is found at the extreme low corner (0,0). This implies that there is a greater probability of a high yield/high yield or extreme low yield/low yield event occurring, with the greatest probability for an extreme low yield/low yield event.

The Gumbel-Hougaard copula places a greater probability on extreme high/high events occurring. In regard to the corn yield data, this means that if an extreme

high corn yield is observed in one county, there is a considerably greater probability that an extreme high corn yield will also be observed in another county within that state, compared to the probability that a low yield will be observed. The Gumbel-Hougaard density plot shows that the height of the density increases towards the low corner (0,0), although the greatest height is found at the extreme high corner (1,1). This implies that there is a greater probability of a low yield/low yield or an extreme high yield/high yield event occurring, with the greatest probability for an extreme high yield/high yield event.

The following tables divide the results by the distance that the corn yields are separated. This is presented to see if the copula selected as having provided the best fit to the data is dependent on the distance that the corn yields are separated. The distance between the corn yields is in the first row of the table and the bivariate copula names (abbreviated) are in the first column of the table[73].

[73] "AMH" represents the Ali-Mikhail-Haq copula, "C-J" represents the Cook-Johnson copula, "F" represent the Frank copula, "G-H" represents the Gumbel-Hougaard copula, "J" represents the Joe copula, "N" represents the Normal copula, and "P" represents the Plankett copula.

Table 22: Copula Results by Distance: Illinois (Absolute)

	0-50	51-100	101-150	151-200	201-250	251-300	301-350	351-400
AMH	0	1	14	46	60	41	11	0
C-J	133	254	281	244	115	59	16	1
F	49	114	195	191	105	41	5	2
G-H	31	117	97	47	21	13	13	2
J	20	106	154	86	50	25	12	4
N	256	515	420	332	215	131	48	8
P	47	119	121	92	53	15	19	4

Table 23: Copula Results by Distance: Illinois (Percentage)

	0-50	51-100	101-150	151-200	201-250	251-300	301-350	351-400
AMH	0.0	0.1	1.1	4.5	9.7	12.6	8.9	0.0
C-J	24.8	20.7	21.9	22.0	18.6	18.2	12.9	4.8
F	9.1	9.3	15.2	18.8	17.0	12.6	4.0	9.5
G-H	5.8	9.5	7.6	4.6	3.4	4.0	10.5	9.5
J	3.7	8.6	12.0	8.4	8.1	7.7	9.7	19.0
N	47.8	42.0	32.8	32.6	34.7	40.3	38.7	38.1
P	8.8	9.7	9.4	9.0	8.6	4.6	15.3	19.0

Table 24: Copula Results by Distance: Indiana (Absolute)

	0-50	51-100	101-150	151-200	201-250	251-300
AMH	0	7	26	46	41	6
C-J	218	484	462	191	29	4
F	118	182	157	70	35	3
G-H	22	41	28	21	15	2
J	29	84	48	49	25	4
N	231	470	392	168	54	4
P	57	116	109	79	52	7

Table 25: Copula Results by Distance: Indiana (Percentage)

	0-50	51-100	101-150	151-200	201-250	251-300
AMH	0.0	0.5	2.1	7.4	16.3	20.0
C-J	32.3	35.0	37.8	30.6	11.6	13.3
F	17.5	13.2	12.8	11.2	13.9	10.0
G-H	3.3	3.0	2.3	3.4	6.0	6.7
J	4.3	6.1	3.9	7.9	10.0	13.3
N	34.2	34.0	32.1	26.9	21.5	13.3
P	8.4	8.4	8.9	12.7	20.7	23.3

Table 26: Copula Results by Distance: Iowa (Absolute)

	0-50	51-100	101-150	151-200	201-250	251-300	301-350
AMH	0	2	5	12	10	4	0
C-J	49	87	39	59	19	9	3
F	38	73	59	63	57	20	0
G-H	91	300	441	318	98	11	0
J	20	61	95	99	34	4	0
N	259	577	518	343	203	68	2
P	72	171	237	153	60	8	0

Table 27: Copula Results by Distance: Iowa (Percentage)

	0-50	51-100	101-150	151-200	201-250	251-300	301-350
AMH	0.0	0.2	0.4	1.1	2.1	3.2	0.0
C-J	9.3	6.8	2.8	5.6	4.0	7.3	60.0
F	7.2	5.7	4.2	6.0	11.9	16.1	0.0
G-H	17.2	23.6	31.6	30.4	20.4	8.9	0.0
J	3.8	4.8	6.8	9.5	7.1	3.2	0.0
N	49.0	45.4	37.2	32.8	42.2	54.8	40.0
P	13.6	13.5	17.0	14.6	12.5	6.5	0.0

Table 28: Copula Results by Distance: North Carolina (Absolute)

	0-50	51-100	101-150	151-200	201-250	251-300	301-350	351-500
AMH	4	12	27	49	57	54	31	32
C-J	96	215	190	120	109	74	49	41
F	62	125	129	106	87	43	40	31
G-H	51	171	137	92	56	55	35	19
J	48	155	192	194	113	78	40	22
N	214	304	263	175	139	78	45	13
P	56	153	105	74	44	24	9	13

Table 29: Copula Results by Distance: North Carolina (Percentage)

	0-50	51-100	101-150	151-200	201-250	251-300	301-350	351-500
AMH	0.8	1.1	2.6	6.0	9.4	13.3	12.4	18.7
C-J	18.1	18.9	18.2	14.8	18.0	18.2	19.7	24.0
F	11.7	11.0	12.4	13.1	14.4	10.6	16.1	18.1
G-H	9.6	15.1	13.1	11.4	9.3	13.5	14.1	11.1
J	9.0	13.7	18.4	24.0	18.7	19.2	16.1	12.9
N	40.3	26.8	25.2	21.6	23.0	19.2	18.1	7.6
P	10.5	13.5	10.1	9.1	7.3	5.9	3.6	7.6

The above tables do not provide any clear evidence that there is a relationship between the best copula and the distance that the corn yields are separated, with the exception of the Ali-Mikhail-Haq copula. The Ali-Mikhail-Haq copula is selected more often as the distance between the corn yields increases, however, overall the Ali-Mikhail-Haq copula is not selected often relative to the other copulas.

Recall that the data set used in this analysis is a relatively small sample of 51 corn yield observations. If a set of corn yield observations contains a few (possibly only one) extreme low yield/low yield events, this could result in the Cook-Johnson copula being selected as the "best" copula. Further, if a set of corn yield observations contains a few (possibly only one) extreme high yield/high yield events, this could result in the Gumbel-Hougaard copula being selected as the "best" copula, and if

a set of corn yield observations contains a few (possibly only one pair) of extreme high yield/high yield and low yield/low yield events, this could result in the Normal copula being selected. Finally, if a set of corn yield observations contains no extreme events, this could result in the Frank or Plankett copulas being selected.

To consider that the results presented above could be a result of a small sample phenomenon, the mean and median odds ratio were calculated for each state. The odds ratio uses the likelihood value of the "best" copula as the numerator and the likelihood value of the "second-best" copula as the denominator. The ratio can be interpreted as the "odds" that the best copula is truly the better of the two copulas. In this context, an odds ratio that is equal to one indicates that there are one-to-one odds that the second-best copula fits the yield data equally as well as the best copula, or it is not possible to distinguish between the two copulas in terms of which copula best fit the data. An odds ratio greater than one[74], for example, an odds ratio of 3, can be interpreted as meaning that there are 3-to-1 odds that the best copula is truly the better of the two copulas. The mean and median odds ratio each of the states is presented below[75].

Table 30: "Best" to "Second-Best" Odds Ratios: Corn Yields

State	**Mean Odds Ratio**	**Median Odds Ratio**
Illinois	3.3	1.5
Indiana	4.3	1.7
Iowa	3.0	1.7
North Carolina	2.7	1.4

For each state, the median odds ratio is less than 2, indicating that there are

[74]The range of values for the odds ratio is $[1, \infty)$.

[75]Because there were a few extreme high odds ratio values, the median odds ratio is presented as well as the mean odds ratio.

greater than 2-to-1 odds that the best copula is truly the better of the two copulas. This result implies that it may be appropriate to use one bivariate copula distribution to describe the bivariate relationship of corn yields. More specifically, it would be of great benefit if the Normal bivariate copula could be used to describe the relationship of corn yields over a particular state. Recall that the multivariate Normal copula distribution is a tractable multivariate copula such that the bivariate copula can easily be expanded to incorporate additional county-level corn yields. This means that if it is appropriate to use the Normal bivariate copula to describe the relationship of corn yields over a particular state, this copula can be expanded to incorporate additional county corn yields and the result would be a multivariate joint distribution of corn yields across a state.

To explore the appropriateness of using the Normal bivariate copula, the mean and median odds ratio for each state is shown below. The odds ratio uses the likelihood value of the "best" copula as the numerator and the likelihood value of the Normal copula as the denominator.

Table 31: "Best" to Normal Odds Ratios: Corn Yields

State	**Mean Odds Ratio**	**Median Odds Ratio**
Illinois	13.6	3.6
Indiana	13.5	4.1
Iowa	12.8	3.5
North Carolina	7.0	2.2

For Illinois, Indiana, and Iowa, the median odds ratio is approximately 4, indicating that there are approximately 4-to-1 odds that the best copula is truly better than the Normal copula. North Carolina's median odds ratio is approximately equal to 2, indicating that there are approximately 2-to-1 odds that the best copula is truly

the better than the Normal copula.

However, it is unclear if 2-to-1 odds, 3-to-1 odds, or even 4-to-1 odds indicate that it is, or is not appropriate to use the Normal copula (or any other copula) to describe the relationship of corn yields over a particular area. A simulation was performed to gain perspective on how close to one the odds ratio needs to be to provide an indication that the second-best copula, or the Normal copula, fits the yield data equally as well as the best copula. The simulation was performed as follows:

1. Two vectors of 51 observations[76], denoted u and v, were randomly drawn from the multivariate normal distribution N(0,Σ), where $\Sigma = \begin{bmatrix} 1 & \varrho \\ \varrho & 1 \end{bmatrix}$ and ϱ is a random uniform value in the range [0,1).

2. The u and v vectors were used to estimate each of the bivariate copulas using maximum likelihood estimation such that:

$$ln\pounds_f = \Sigma_{i=1}^{n} ln[c(u_i, v_i \mid \hat{\beta}]$$

where $\pounds_f$ represents the likelihood function for the bivariate copula density f, $\hat{\beta}$ is/are the unknown copula parameters to be estimated by finding the $\hat{\beta}$ that maximizes $\pounds_f$, and n is equal to 51.

3. The odds ratio for the best copula to the second-best copula and the odds ratio for the best copula to the Normal copula were calculated and retained.

4. The above steps were repeated 10,000 times.

[76]The 51 observations correspond to the 51 crop yield observations contained in the data set used in this analysis.

5. The 2.5 and 97.5 percentile of both sets of the 10,000 odds ratios were calculated.

The results are as follows:

Table 32: Odds Ratio Simulation Results

Odds Ratio	**2.5 Percentile**	**97.5 Percentile**
Best to Second-Best	1.0148	84.4
Best to Normal	1.0000	607.5

These simulation results suggest that the mean and median odds ratios calculated from the corn yield data are not outliers such that their values do not fall in the tails of the odds ratio distribution. In other words, the mean and median odds ratios calculated from the yield data are not unlikely values such that they are not exceptionally low or exceptionally high values. Although the simulation results suggest that the odds ratio values calculated from the corn yield data are not extreme values, these results do not indicate the appropriateness of using the Normal copula (or any other copula) to describe the relationship of corn yields over a particular area. Perhaps the best way to consider the appropriateness of using one copula to describe the relationship of crop yields is to consider how much it would cost an insurer if the "second-best" copula, or the Normal copula is used to calculate insurance premiums when, in actuality, another copula best fit the crop yield data. This issue to addressed in the section "Copula Methodology: Crop Insurance Application".

0.5.2 Application to Corn and Soybean Yields

The copula methodology was applied to the county-level corn and soybean yield data, as described in the "Data" section, for each of the four states considered in this analysis. The application to the corn and soybean yield data is the same as described in the section "Application to Corn Yields", with the exception that the corn and soybean yields from the same county are used rather than the corn yields from different counties. The results are as follows:

Table 33: Copula Results: Corn and Soybean Yields (Absolute)

Bivariate Copula	**Illinois**	**Indiana**	**Iowa**	**North Carolina**
Ali-Mikhail-Haq	0	1	0	13
Cook-Johnson	43	49	52	24
Frank	10	3	11	7
Gumbel-Hougaard	11	0	0	13
Joe	2	4	0	6
Normal	25	27	35	15
Plackett	11	8	1	9

Table 34: Copula Results: Corn and Soybean Yields (Percentage)

Bivariate Copula	**Illinois**	**Indiana**	**Iowa**	**North Carolina**
Ali-Mikhail-Haq	0.0	1.1	0.0	14.9
Cook-Johnson	42.2	53.3	52.5	27.6
Frank	9.8	3.3	11.1	8.0
Gumbel-Hougaard	10.8	0.0	0.0	14.9
Joe	2.0	4.3	0.0	6.9
Normal	24.5	29.3	35.4	17.2
Plackett	10.8	8.7	0.0	10.3

The two bivariate copulas that are most often selected as the best fitting copula are the Normal and Cook-Johnson copulas. These results are similar to the corn yield copula results and the characteristics of these copulas are discussed in the previous section "Application to Corn Yields".

Again, to consider that the results presented above could be a result of a small sample phenomenon, the mean and median odds ratio were calculated for each state. Similar to the odds ratio for the "best" copula to the "second-best" copula, presented in the previous section, the odds ratio presented below uses the likelihood value of the best copula as the numerator and the likelihood value of the second-best copula as the denominator.

Table 35: "Best" to "Second-Best" Odds: Corn and Soybean Yields

State	**Mean Odds Ratio**	**Median Odds Ratio**
Illinois	5.3	2.0
Indiana	3.5	2.1
Iowa	4.5	3.2
North Carolina	1.8	1.3

For Illinois and Indiana, the median odds ratio is approximately equal to 2, indicating that there are approximately 2-to-1 odds that the best copula is truly the better of the two copulas. The median odds ratio for Iowa is slightly higher at 3.2 and the median odds ratio for North Carolina is slightly lower at 1.3.

To explore the appropriateness of using the Normal copula to describe the relationship of corn and soybean yields, the mean and median odds ratio for the best copula to the Normal copula for each state is shown below. The odds ratio uses the likelihood value of the "best" copula as the numerator and the likelihood value of the Normal copula as the denominator.

Table 36: "Best" to Normal Odds: Corn and Soybean Yields

State	Mean Odds Ratio	Median Odds Ratio
Illinois	15.7	5.5
Indiana	11.4	5.3
Iowa	28.5	10.4
North Carolina	4.3	2.1

For Illinois and Indiana, the median odds ratio is approximately equal to 5, indicating that there are greater than 5-to-1 odds that the best copula is truly the better of the two copulas. The median odds ratio for Iowa is considerably higher at 10.4 and the median odds ratio for North Carolina is considerably lower at 2.1.

The odds ratio simulation described in the previous section suggests that the mean and median odds ratios calculated from the corn and soybean yield data are not outliers such that their values do not fall in the tails of the odds ratio distribution. In other words, the mean and median odds ratios calculated from the yield data are not unlikely values such that they are not exceptionally low or exceptionally high values. Although the simulation results suggest that the odds ratio values calculated from the corn and soybean yield data are not extreme values, these results do not indicate the appropriateness of using the Normal copula (or any other copula) to describe the relationship of corn and soybean yields over a particular area. The following section "Copula Methodology: Crop Insurance Application" considers how much it would cost an insurer if the "second-best" copula, or the Normal copula is used to calculate insurance premiums when, in actuality, another copula best fit the crop yield data. This is perhaps the best method for weighing the benefits of using one copula, the Normal copula in particular, to describe the relationship of crop yields to the potential costs of doing so.

0.5.3 Copula Methodology: Crop Insurance Application

This section demonstrates how the copula methodology can be applied to the crop insurance industry. More specifically, this analysis compares the probabilities that an indemnity payment is made and the expected indemnity payments implied by each of the bivariate copulas for a representative county's corn and soybean yields. The odds ratio results from the previous section gave the likelihood that the "best" copula truly provided a better fit to the yield data compared to either the "second-best" copula or the Normal copula. However, the odds ratio results do not provide enough information to determine if it is appropriate to use one copula, the Normal copula in particular, to describe the relationship of crop yields. The copula density plots shown previously had considerably different shapes, indicating that the probability that an indemnity is paid and the expected indemnity payment may be significantly different across the copulas. In order to properly weigh the pros and cons of using one copula, the Normal copula in particular, to describe the relationship of crop yields, the potential cost of not using the "best" copula must be determined. The following analysis explores these potential costs by using the crop yield data from a county selected at random from the data set.

The corn and soybean yield data from Story county Iowa are used to compare the probabilities that an indemnity payment is made and the expected indemnity payments implied by each of the copulas for a whole-farm revenue insurance policy. The Spearman rank correlation coefficient between these two crop yields is 0.77. In order to continue with this analysis, the parameters of the crop insurance policy need to be set and the crop yield data needs to be fit to a distribution. The following crop

insurance policy parameters were assumed[77]:

Table 37: Crop Insurance Policy Parameters

Parameter	**Value**
Price of Corn per Bushel	\$2.40
Price of Soybeans per Bushel	\$4.80
Acres of Corn Insured	100
Acres of Soybeans Insured	100
Election Level	75%

The corn and soybean yield data were fit to Beta(p,q) distributions. Appendix B (Part Two) provides a discussion of the method used to fit the yield data to the Beta(p,q) distributions. The following table summarizes these results:

Table 38: Beta($\boldsymbol{p}, \boldsymbol{q}$): Story County Corn and Soybeans

Crop	$\boldsymbol{p}$	$\boldsymbol{q}$	**Beta(p, q) Mean Value**
Corn	5.933	1.888	0.7585
Soybeans	3.979	1.466	0.7307

The indemnity function for these parameters is defined as:

$$I(c, s) = 100\left[(\$2.40 * (.75 * \mu_c - x_c)) + (\$4.80 * (.75 * \mu_s - x_s))\right] \tag{4}$$

where μ_c is the inverse Beta(p,q) CDF of the mean of the corn yields which has been converted to "yield" terms[78], μ_s is the inverse Beta(p,q) CDF of the mean of the soybean yields which has been converted to "yield" terms, x_c is the observed value of the corn yield (in "yield" terms) , and x_s is the observed value of the soybean yield (in "yield" terms).

[77]The election level is the percentage of the mean crop yield that is to be insured. An election level of 75% indicates that the observed crop yield must fall below 75% of its mean value in order for an indemnity to be paid.

[78]Converted to "yield" terms means that the Beta value has been converted back to its implied yield value using the following equation: $\mu_c = \tilde{\mu}_c(b_c - a_c) + a_c$, where $\tilde{\mu_c}$ is the Beta(p,q) value of the mean of the corn yield, a_c is the minimum value of the corn yield, and b_c is the maximum of the corn yield. See Appendix B (Part Two) for further discussion.

Given the above parameters of the yield distributions, the correlation between the yields, and the crop insurance policy, a comparison across the copulas of the probability that an indemnity is paid can be computed. The following figure is presented below to aid in this discussion. The figure shows a representative copula distribution on the unit square. The distribution is divided into four sections using $\hat{c}$ and $\hat{s}$, where $\hat{c}$ is equal to 0.1125, which is the Beta(p,q) CDF value of 75% of the mean value of the corn yields (β(0.5688, 5.933, 1.888), where 0.5688 = .75*0.7585) and $\hat{s}$ is equal to 0.1619 which is the Beta(p,q) CDF value of 75% of the mean value of the soybean yields (β(0.5480, 3.979, 1.466), where 0.5480 = .75*0.7307)[79]. Area C represents the area of the copula distribution in which an indemnity is paid because the observed corn yield falls below $\hat{c}$, that is, the area of the distribution where $x_c < \hat{c}$ and $x_s > \hat{s}$. Area A represents the area of the copula distribution in which an indemnity is paid because the observed soybean yield falls below $\hat{s}$, that is, the area of the distribution where $x_s < \hat{s}$ and $x_c > \hat{c}$. Area B represents the area of the copula distribution in which an indemnity is paid because the observed corn yield falls below $\hat{c}$ and the observed soybean yield falls below $\hat{s}$, that is, the area of the distribution where $x_c < \hat{c}$ and $x_s < \hat{s}$. Area D represents the area of the distribution where an indemnity is not paid because the observed corn yield is greater than or equal to $\hat{c}$ and the observed soybean yield is greater than or equal to $\hat{s}$, that is, the area of the distribution where $x_c \geq \hat{c}$ and $x_s \geq \hat{s}$.

[79]Recall that an indemnity is paid if the observed corn yield, soybean yield, or both yields fall below 75% of their mean value.

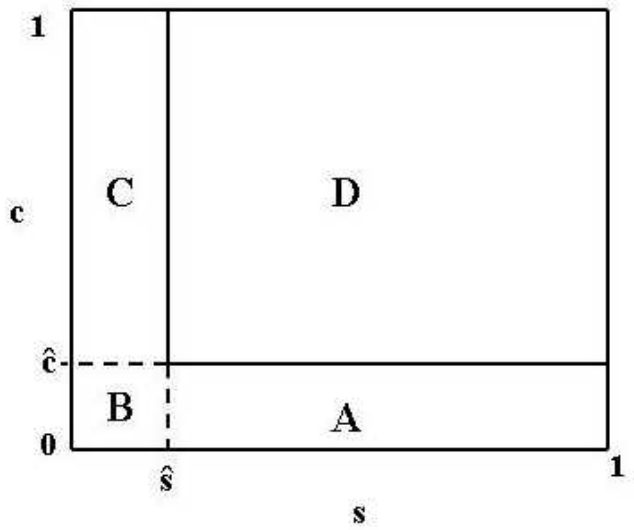


Figure 43: Representative Copula Distribution

Given the same underlying marginal distributions and the same Spearman correlation between yields, each copula implies different probabilities for areas A, B, C, and D. To demonstrate this, each of the copulas distributions were evaluated over the areas A, B, C, and D using $\hat{c}$, $\hat{s}$, and the Spearman correlation calculated for the Story county corn and soybean yields[80]. The results are shown in the following table. The table lists the copula and the probability of an event occurring in each of the areas A, B, C, D, and I, where I is the probability that an indemnity is paid[81].

[80] $Pr(A) = \hat{c} - C(\hat{c}, \hat{s})$ $Pr(B) = C(\hat{c}, \hat{s})$ $Pr(C) = \hat{s} - C(\hat{c}, \hat{s})$ $Pr(D) = 1 - Pr(A) - Pr(B) - PR(C)$

where C(.) is the copula CDF, Pr(A) is the probability that an observation occurs in area A, Pr(B) is the probability that an observation occurs in area B, and so on.

[81] I = A + B + C

Table 39: Probability of Indemnity: Story County Corn and Soybeans

Copula	**A**	**B**	**C**	**D**	**I**
Ali-Mikhail-Haq	0.0918	0.0701	0.0424	0.7957	0.2043
Cook-Johnson	0.0612	0.1007	0.0118	0.8263	0.1737
Frank	0.0953	0.0665	0.0461	0.7921	0.2079
Gumbel-Hougaard	0.0951	0.0668	0.0457	0.7924	0.2076
Joe	0.0714	0.0904	0.0221	0.8161	0.1839
Normal	0.0851	0.0768	0.0357	0.8024	0.1976
Plackett	0.0861	0.0759	0.0366	0.8014	0.1986

For the parameters described above, the Frank copula has the greatest probability that an indemnity payment is made either because the observed corn yield, observed soybean yield, or both yields fall below $\hat{c}$ or $\hat{s}$, respectively. The Cook-Johnson copula has the smallest probability that an indemnity payment is made, although it has the greatest probability for both the observed corn yield and the observed soybean yield to fall below $\hat{c}$ or $\hat{s}$, respectively. The probabilities of an event occurring in the different areas of the Normal copula tend to be in the middle range compared to the other copulas.

Finally, a comparison of the expected indemnity payments, both in dollar terms and in percentage of total loss, is made. For each of the copulas, the indemnity function defined in Eq. (4) was evaluated over the regions A, B, and C using the following procedure:

1. A bivariate set of 50,000 quasi-random observations were drawn over the areas A, B, and C.

2. Each of the copulas were evaluated at each of the 50,000 pairs of observations to obtain 50,000 indemnity payments.

3. The mean and standard deviation of the indemnity payments implied by each of the copulas were calculated.

The cost to the insurer of the total loss, that is, a zero yield for both corn and soybeans, for the indemnity function described in Eq. (4) was also calculated[82]. For each copula, the percentage that the mean indemnity payment is to the total loss was calculated by dividing the mean indemnity payment by the total loss and multiplying by 100. The results are as follows:

Table 40: Indemnity Payments

Copula	**Mean**	**St. Deviation**	**% of Total Loss**
Ali-Mikhail-Haq	\$5525.61	\$751.82	9.50%
Cook-Johnson	\$5042.44	\$619.35	8.67%
Frank	\$5583.12	\$767.64	9.60%
Gumbel-Hougaard	\$5578.73	\$766.41	9.59%
Joe	\$5218.77	\$798.12	8.97%
Normal	\$5417.15	\$722.62	9.31%
Plackett	\$5431.46	\$726.58	9.33%

As expected, the Frank copula has the greatest mean or expected indemnity payment and this indemnity payment is the greatest percentage of total loss compared to the other copulas. Further, the Cook-Johnson copula has the smallest expected indemnity payment and this indemnity payment is the smallest percentage of total loss compared to the other copulas. The expected indemnity payment, and its percentage of total loss, for the Normal copula falls near the middle range compared to the other copulas.

When interpreting these results, it is important to keep in mind that these results are for only 200 total insured acres from one county. Realistically, a crop insurer in-

[82]The cost to the insurer of a total loss is \$58,152.78.

sures considerably more than 200 acres. For the 200 acre case, the difference between the percentage of total loss implied by the Gumbel-Hougaard copula (9.59%) and the percentage of total loss implied by the Normal copula (9.31%), or approximately $160 difference in expected indemnity payments, may not seem significant. However, if one were to consider 50,000 total insured acres instead of 200, the difference in expected indemnity payments will be significantly greater.

Further, consider the odds ratios for Iowa. The median odds ratio of the "best" copula to the Normal copula for the corn and soybean yield data is 10.4. This indicates that there are approximately 10-to-1 odds that the best copula truly fit the yield data better than the Normal copula. If the "best" copula is the Gumbel-Hougaard copula, which occurred frequently in the Iowa data, and the Normal copula is used to calculate the expected indemnity payment and insurance premiums, the likelihood that the expected indemnity payment is under-estimated and the resulting insurance premiums are below the actuarial fair rate is significant. Again, if one considers the situation in which several thousand acres are insured, using the Normal copula instead of the Gumbel-Hougaard copula will cause the insurer long-run financial loss as the insurance premium will be set below the actuarial fair rate[83].

Finally, consider the situation in which the Cook-Johnson copula is the best copula, but the Normal copula is used instead to determine the expected indemnity payment and insurance premiums. In this situation, the expected indemnity payment will be over-estimated resulting in insurance premiums that are above the actuarial fair rate. As discussed in Goodwin[7], insurance premiums set too high are likely to

[83]This assumes that the Gumbel-Hougaard copula is truly the "best" copula.

exacerbate the adverse selection problem which will lead to long-run financial loss for crop insurers[84].

The above results indicate that care needs to be taken when selecting a copula to describe the relationship of crop yields. The copula that is selected to describe the relationship of crop yields is ultimately used to determine the crop insurance premiums. If an inappropriate copula is selected and used to determine insurance premiums, the insurer is likely to suffer long-run financial loss as a result of setting the insurance premiums above or below the actuarial fair premium rate.

[84]The problem of adverse selection is discussed in the section "Causes of the FCIC's Financial Loss" in this paper.

0.6 Research Summary and Conclusions

Systemic risk, in regard to agricultural production, refers to the spatial dependence of crop yields stemming from correlated weather, soil patterns, and other geographically related factors. Systemic risk has been named as a contributing factor to the Federal Crop Insurance Corporation's (FCIC) poor actuarial performance. To date, little research has explored the spatial dependence structure of crop yields. This paper provided an in-depth study of the spatial dependence structure of crop yields and considered the implications for crop insurance. The following provides a summary of the three main analyses in this paper; the equality of the cross-crop correlation, the rate of spatial correlation decay, and the bivariate copula distributions that describe the spatial dependence structure of crop yields.

In Section 0.3, the equality of the county-level cross-crop correlation for corn and soybean yields was tested in each of the states; Illinois, Indiana, Iowa, and North Carolina. The county-level cross-crop yield correlation $\rho_{x,mn}$ is the correlation between crop yield m and crop yield n in county x. If t crops are produced in county x, the cross-crop yield correlation matrix Ω_x for county x is a (t x t) matrix with ones down the diagonal and $\rho_{x,mn}$ is the m, n^{th} element of Ω_x. The focus of this analysis is to determine if the cross-crop yield correlation $\rho_{x,mn}$ drifts slowly enough across space such that it can be considered the same over a particular geographical area, in this case, a state.

The cross-crop yield correlation is of particular interest to whole-farm revenue insurers. Whole-farm revenue insurance offers farmers protection that is based on the "whole-farm's" revenue. An insured farmer receives an indemnity payment if the

total revenue earned by his farm that period falls below a predetermined level. The cross-crop correlation between the insured farmer's crop yields affects the portfolio risk faced by the whole-farm revenue insurer. In assessing the portfolio risk faced by a whole-farm revenue insurer and in calculating insurance premiums for whole-farm revenue insurance policies, the question arises as to how the cross-crop correlation should be calculated. The answer largely depends on whether the cross-crop correlation is constant across space, or if it is location-dependent.

Two nonparametric tests, the Sign and Friedman tests, were used to test the equality of the cross-crop correlation parameter across space using the county-level corn and soybean yield from Illinois, Indiana, Iowa, and North Carolina. The cross-crop correlation is found to be location-dependent in Illinois, but not location-dependent in the other states. Three alternative methods are suggested for calculating the cross-crop correlation in those states in which the correlation is not location-dependent. These methods calculate the cross-crop correlation parameter by pooling the county-level data or by using state-level data. The cross-crop correlation estimate was calculated using each of the three alternate methods using the corn and soybean yield data from Indiana, Iowa, and North Carolina[85]. Compared to calculating the cross-crop correlation by county, these alternative methods can lead to more efficient estimates of the cross-crop correlation, as calculating the cross-crop correlation for a particular county relies solely on the availability of that county's data. Which method is optimal will depend on the availability of county-level and state-level data, researcher preferences and *a priori* beliefs, among other consider-

[85]The cross-crop correlation was found not be location dependent in each of these states.

ations. If there are a significant number of missing county-level observations, the methods described and used in this analysis may provide better estimates of the cross-crop correlation when the cross-crop correlation is found not to be location dependent.

Section 0.4 modeled the drift rate of the correlation parameter between same crop yields through space, or more specifically, the rate of spatial correlation decay of crop yields. The rate of spatial correlation decay of crop yields is the rate at which the correlation between counties' crop yields decreases as the distance between counties increases. This decay rate provides two main insights for crop insurers. It provides an indication of the distance between crop yields that is required for the yields to be no longer significantly correlated, and an indication of degree to which the spatial correlation of crop yields has impacted the FCIC's actuarial performance.

Several rate of spatial correlation decay models were developed and tested. The following model:

$$\rho_{xy,c} = exp(\beta_1 d_{xy} + \beta_2 d_{xy}^2 + \beta_3 d_{xy}^3)\nu_{xy,c}$$

provided the best overall fit to the county-level corn yield data from the four states considered in this analysis. Several properties of this model were explored. The sign definiteness of the variance/covariance matrix (in distance) of this model was examined. Although a proof of the sign definiteness of this matrix is beyond the scope of this paper, the simulation study performed indicated that the variance/covariance matrix is likely to be positive definite in distance. The property of monotoncity was also examined. For each of the states considered, the estimated rate of spatial correlation decay model is strictly monotonically decreasing.

The above model was extended to allow for the rate of spatial correlation decay model to depend on the direction as well as the distance that the crop yields are separated. The rate of spatial correlation decay model was found to be dependent on direction as well as the distance that crop yields are separated for the Iowa and North Carolina corn yield data. The estimation results were presented and interpreted graphically, as well as in tabular form. The results of these analyses indicated that both the Iowa and North Carolina models are strictly monotonically decreasing in distance. Further, a simulation indicated that the variance/covariance matrix for the Iowa model was not positive definite in distance, but the North Carolina model is likely to be positive definite in distance.

There are many implications for crop insurers that can be drawn from the rate of spatial correlation decay analysis. Consider the rate of spatial correlation graphs (Figures 22-25). The county-level crop yields from the states considered in this analysis are significantly correlated to distances as great as 475 miles. Therefore, the size of the geographical area that crop insurers need to be concerned about the spatial correlation of yields is at least a 475 mile radius. Further, because the geographical area that contains correlated crop yields is relatively large, this indicates that the spatial correlation of crop yields is likely to be a significant contributor to the FCIC's poor actuarial performance.

Section 0.5, the final analysis section, explored the bivariate distributions of crop yields using the copula methodology. These bivariate distributions allowed for an exploration of the higher-order spatial relationships, or higher-order cross-moments of crop yields. Further, these bivariate distributions can be used to develop crop

insurance premiums that account for the spatial correlation of crop yields. The purpose of this analysis is to determine the "best" bivariate distribution for crop yields and to determine if it is appropriate to use one bivariate distribution over a geographical area, in particular, the Normal bivariate copula.

The two bivariate copulas that were most often selected as the "best" fitting copula for the corn and soybean yield data were the Cook-Johnson and Normal copulas. The Gumbel-Hougaard copula is frequently selected as the best copula using the Iowa corn yield data only, but it did not occur frequently in the other states. The Normal copula places an approximately equal probability on extreme low/low events and extreme high/high events. In regard to crop yields, this means that if an extreme high yield is observed in one county, there is a considerably greater probability that an extreme high yield will also be observed in another county within that state, compared to the probability that a low yield will be observed. And, if an extreme low yield is observed in one county, there is a considerably greater probability that an extreme low yield will also be observed in another county in that state, compared to the probability that a high yield will be observed. Further, the Normal copula plot indicates that there is a very low probability of a high yield/low yield or low yield/high yield event occurring compared to the probability of a high yield/high yield or low yield/low yield event occurring. The Cook-Johnson copula places a greater probability on extreme low/low events occurring. In regard to the crop yields, this means that if an extreme low yield is observed in one county, there is a considerably greater probability that an extreme low yield will also be observed in another county within that state, compared to the probability that a high yield will

be observed. Further, the Cook-Johnson density plot shows that there is a greater probability of a high yield/high yield or extreme low yield/low yield event occurring, with the greatest probability for an extreme low yield/low yield event.

An application of the copula methodology to a representative county's yield data and insurance policy was performed. In this application, the probabilities that an indemnity payment is made were compared across each of the copulas. The Frank copula yielded the greatest probability for an indemnity payment and the Cook-Johnson copula yielded the smallest probability for an indemnity payment. The expected indemnity payment and the percentage that the expected indemnity payment was of the total loss were also compared across the copulas. As expected, the Frank copula implied the greatest expected indemnity payment which was the greatest percentage of total loss and the Cook-Johnson copula implied the smallest expected indemnity payment which was the smallest percentage of total loss.

The results of this analysis section indicated that care needs to be taken when selecting a copula to describe the relationship of crop yields. The copula that is selected to describe the relationship of crop yields is ultimately used to determine the crop insurance premiums. If an inappropriate copula is selected and used to determine insurance premiums, two different situations could arise, each with the same result for the insurer. If the copula distribution selected to determine crop insurance premiums underestimates the expected indemnity payment, the insurer is likely to suffer long-run financial loss as a result of setting the insurance premiums below the actuarial fair premium rate. Further, if the copula distribution selected to determine crop insurance premiums overestimates the expected indemnity payment

and the premium rate is set above the actuarial fair rate, the insurer is likely to suffer long-run financial loss because the insurance pool is likely to become more adversely selected.

Overall, the analyses performed in this paper provided insights into several aspects of the spatial dependence structure of crop yields. Given these insights and the results of the analyses, alternative crop insurance practices and were offered. First, in some states, it may be appropriate to use pooling methods or state-level data to calculate the cross-crop correlation for use in whole-farm revenue insurance policies. Given that county-level data often contains missing data, these alternative methods for calculating the cross-crop correlation parameter can provide better estimates of the cross-crop correlation. Second, crop insurers need to consider that yields separated by several hundred miles may be significantly correlated and this issue needs to be accounted for in the calculation of expected indemnity payments and insurance premiums. Finally, the copula methodology is suggested as a method for estimating the bivariate distributions of crop yields, as these distributions will account for the spatial dependence of crop yields. These bivariate distributions can be used to calculate expected indemnity payments and insurance premiums which will, in turn, account for the spatial dependence of crop yields. However, care must be taken in selecting these copula distributions, for if an inappropriate distribution is selected, this can result in long-run financial loss.

Bibliography

[1] Conover, W.J. Practical Nonparametric Statistics, 2nd Edition. New York: John Wiley & Sons, Inc., 1971.

[2] Frees, E.W. and E.A. Valdez. "Understanding Relationships Using Copulas." *North American Actuarial Journal.* Vol. 2, Number 1 1997: 1-25.

[3] Gallagher, P. "U.S. Corn Yield Capacity and Probability: Estimation and Forecasting with Nonsymmetric Disturbances." *N.C. Journal of Agricultural Economics.* Vol.8, Jan 1986: 109-22.

[4] Gallagher, P. "U.S. Soybean Yields: Estimation and Forecasting with Nonsymmetric Disturbances." *American Journal of Agricultural Economics.* Vol. 69, Nov 1987: 796-803.

[5] General Acconting Office Report, Number RECD-98-111. "Problems With New Crop Revenue Insurance Plans."

[6] General Acconting Office Report, Number RECD-99-161. "Crop Insurance: Further Actions Could Strenthen Program's Financial Soundness."

[7] Goodwin, B.K. "An Empircial Analysis of the Demand for Multiple Peril Crop Insurance," *American Journal of Agricultural Econmics.* Vol. 75, May 1993: 425-434.

[8] Goodwin, B.K. "Problems With Market Insurance in Agricultural." Unpublished paper. Jan. 2001.

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

[10] Goodwin, B.K. and V.H. Smith. The Economics of Crop Insurance and Disaster Aid. Washington, D.C.: AEI Press, 1995.

[11] Haas, C.N. "On Modeling COrrelated Random Variables in Risk Assessment," *Risk Analysis.* Vol. 19, No. 6 1999: 1205-1213.

[12] Hansen, B.E. "Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis," *Economtrica.* Vol. 64, 1996: 413-430.

[13] Hansen, B.E. "Inference In TAR Models," *Studies in Nonlinear Dynamics and Econometrics.* MIT Press, Vol.2, Number 1, April 1997: 1-14.

[14] Joag-Dev, K. "Measures of Dependence." P.R. Krishnaiah and P.K. Sen, eds.*Handbook of Statistics, Vol. 4* New York, Elsevier Science Publishers, 1984: 78-88.

[15] Kennedy, P. A Guide to Econometrics. Cambridge, Massachusetts: The MIT Press, 1992.

[16] Lambert, R. "Long-Term Contracts and Moral Hazard." *Bell Journal of Economics.* Vol. 14, Autumn 1983: 441-52

[17] Miranda, M.J. "Area-Yield Crop Insurance Reconsidered," *American Journal of Agricultural Econmics.* Vol. 73, May 1991: 223-42.

[18] Miranda, M.J. and J.W. Glauber "Systemic Risk, Reinsurance, and the Failure of Crop Insurance Markets." *Amercian Journal of Agricultural Economics.* Vol. 79, Feb. 1997: 206-215.

[19] Nelson, C.H. and P.V. Preckel. "The Conditional Beta Distribution As a Stochastic Production Function." *N.C. Journal of Agricultural Economics.* Vol. 12, Jan 1990: 71-78.

[20] Parzen, E. "On Estimation of a Probability Density and Mode." *Ann. Math. Statist.* Vol. 33, September 1962: 1065-76.

[21] Ramirez, O.A. "Estimation and Use of Multivariate Parametric Models for Simulation of Heteroskedastic, Correlated, Nonnormal Random Variables: The Case of Corn Belt Corn, Soybean, and Wheat Yields." *American Journal of Agricultural Economics.* Vol. 79, Feb. 1997: 191-205.

[22] Rogerson, W. "Repeated Moral Hazard." *Econometrica* Vol. 53, Jan 1985: 69-76.

[23] Schweizer, B., and E.F. Wolff. "On Nonparametric Measures of Dependence for Random Variables." *The Annals of Statistics* Vol. 9, 1981: 879-885.

[24] Silverman, B.W. Density Estimation for Statistics and Data Analysis. London: Chapman & Hall, 1986.

[25] Sklar, A. "Fonctions de reparition a n dimensions et leurs marges," *Publ. Inst. Stat. Univ. Paris* Vol. 8, 1959: 229-231.

[26] Swinton, S. and R.P. King. "Evaluating Robust Regresson Techniques for Detrending Crop Yield Data and Nonnormal Errors." *american Journal of Agricultural Economics.* Vol. 73, May 1991: 446-51.

[27] Vercammen J. and G.C. van Kooten "Moral Hazard Cycles in Individual-Coverage Crop Insurance," *Amercian Journal of Agricultural Economics.* Vol. 73, May 1991: 446-61.

[28] Wackerly, D.D., W. Mendenhall III and R.L. Scheaffer. Mathematical Statistics with Applications, 5th Ed. Belmont: Duxbury Press 1996.

[29] Wang, H. "Zone-Based Group Risk Insurance," *Journal of Agricultural and Resource Economics.* Vol. 25(2), December 2000: 411-31.

0.8 Appendix A: Rate of Spatial Correlation Decay

The following is a list of the other models explored to estimate the rate of spatial correlation decay. The model described in Eq. (2) provided the best overall fit using the criteria described previously[86].

$$\rho_{xy,c} = exp(\beta_1 d_{xy})\nu_{xy,c} \tag{5}$$

$$\rho_{xy,c} = exp(\beta_1 d_{xy} + \beta_2 d_{xy}^2)\nu_{xy,c} \tag{6}$$

$$\rho_{xy,c} = exp(\alpha + \beta_1 d_{xy})\nu_{xy,c} \tag{7}$$

$$\rho_{xy,c} = exp(\alpha + \beta_1 d_{xy} + \beta_2 d_{xy}^2)\nu_{xy,c} \tag{8}$$

$$\rho_{xy,c} = exp(\alpha + \beta_1 d_{xy} + \beta_2 d_{xy}^2 + \beta_3 d_{xy}^3)\nu_{xy,c} \tag{9}$$

$$\rho_{xy,c} = exp(\beta_1 d_{xy} + \beta_2 d_{xy}^2 + \beta_3 d_{xy}^3 + \beta_4 d_{xy}^4)\nu_{xy,c} \tag{10}$$

where the α is an intercept term and the remaining variables and parameters are the same as described previously. For each of the models described above, the null hypothesis of H_o: $\beta = 0$ was failed to be rejected for at least one of the parameter estimates using the t-value, and the LM, LR, and Wald tests, or the overall fit of the model deteriorated (using the criteria described previously) compared to the model described in Eq. (2).

[86]Each of the models described below were fit to all four states' county-level corn yield data. The estimation results for all four states were considered jointly, and the model described in Eq. (2) provided the best overall fit to these four data sets, using the criteria described previously.

0.9 Appendix B: Bivariate Distributions of Crop Yields

0.9.1 Appendix B (Part One)

The following lists the bivariate copula distributions used in this analysis, their parameter restrictions, and, where possible, the relationship between the parameter and Spearman's correlation coefficient. Schweizer and Wolff[23] show that the general relationship between a bivariate copula and Spearman's ρ between two random variables X_1 and X_2 (with a jointly continuous distributions function) is given by:

$$\rho(X_1, X_2) = 12 \int \int (C(u, v) - uv)\, dudv$$

Note that all of the following bivariate copulas are single parameter copulas with the exception of the Joe copula which is a two-parameter copula.

Ali-Mikhail-Haq Copula:

$$C(u, v) = uv[1 - \alpha(1 - u)(1 - v)]^{-1}$$

Parameter Restriction: $-1 \leq \alpha \leq 1$

Relationship to Spearman's ρ: Complicated form

Cook-Johnson Copula:

$$C(u, v) = [u^{-\alpha} + v^{-\alpha} - 1]^{\frac{-1}{\alpha}}$$

Parameter Restriction: $\alpha \geq 0$

Relationship to Spearman's ρ: Complicated form

Frank Copula:

$$C(u,v) = \alpha^{-1} \ln\left(1 + \frac{(\exp^{\alpha u} - 1)(\exp^{\alpha v} - 1)}{(\exp^{\alpha} - 1)}\right)$$

Parameter Restriction: $\alpha \neq 0$

Relationship to Spearman's ρ:

$$\rho = 1 - \frac{12}{\alpha}[D_2(-\alpha) - D_1(-\alpha)]$$

where $D_k(x)$ is the Debye function and, for negative arguments is defined as:

$$D_k(-x) = D_k(x) + \frac{kx}{k+1}$$

where

$$D_k(x) = \frac{k}{x^k} \int_0^x \frac{uv^k}{e^u v - 1} dt$$

for $k = 1, 2$.

Gumbel-Hougaard Copula:

$$C(u,v) = \exp\left[-\left[(-\ln u)^{\alpha} + (-\ln v)^{\alpha}\right]^{\alpha^{-1}}\right]$$

Parameter Restriction: $\alpha \geq 1$

Relationship to Spearman's ρ: No closed form

Joe Copula:

$$C(u,v) = \delta^{-1}\left[1 - \left[1 - \frac{(1-\delta u)^{\theta}(1-(1-\delta v)^{\theta})}{(1-(1-\delta))^{\theta}}\right]^{\theta^{-1}}\right]$$

Parameter Restrictions: $\theta \geq 0$ and $0 \leq \delta \leq 1$

Relationship to Spearman's ρ: One parameter is fixed and a grid search is used to determine the second.

Normal Copula:

$$C(u,v) = G(\Phi^{-1}(u), \Phi^{-1}(v))$$

where G(.) is the bivariate normal distribution function with the correlation coefficient ρ and Φ^{-1} is the inverse of the univariate normal distribution function.

Parameter Restriction: $-1 < \rho < 1$

Relationship to Spearman's ρ:

$$\rho = \frac{6}{\pi}\arcsin(\frac{\alpha}{2})$$

Plackett Copula:

$$C(u,v) = .5(\alpha-1)^{-1}\left[1 + (\alpha-1)(u+v) - [(1+(\alpha-1)(u+v))^2 + 4\alpha(1-\alpha)]^{\frac{1}{2}}\right]$$

Parameter Restriction: $\alpha \geq 0$

Relationship to Spearman's ρ:

$$\rho = \frac{\alpha+1}{\alpha-1} - \frac{2\alpha\ln\alpha}{(\alpha-1)^2}$$

0.9.2 Appendix B (Part Two)

The Story county corn and soybean yields were fit to a Beta(p,q) distribution. Recent research by Nelson and Preckel[19], Gallagher[3], Gallagher[4], Swinton and King[26], and Ramirez[21] has suggested that crop yield distributions are negatively skewed as a result of biological limitations that restrict the maximum yield produced and the common occurrence of negative environmental factors, such as weather and pest damages, that often result in lower yields. The Beta distribution is chosen in this analysis because it allows for the negative skew of crop yields.

The following describes the method used to fit the Story county corn and soybean yield data to Beta distributions:

1. The Story county corn and soybean yields were detrended using the method described in the section "Method Used to Remove Time Trend" and the $\epsilon_{l,c}$ residuals were retained.

2. The residuals were converted to a [0,1] scale using the following formula:

$$z_{l,c} = \frac{\epsilon_{l,c} - a}{b - a}$$

where $z_{l,c}$ is the transformed residual vector, a is the estimated minimum and b is the estimated maximum of the residual vectors. The minimum and maximum values were estimated using the properties of order statistics such that:

$$a = \frac{T\min(\epsilon_{l,c}) - \max(\epsilon_{l,c})}{(T-1)}$$

$$b = \frac{T\max(\epsilon_{l,c}) - \min(\epsilon_{l,c})}{(T-1)}$$

where T is the number of $\epsilon_{l,c}$ observations[87].

3. The method of moments was used to solve for the Beta(p,q) parameters. The first two moments of the Beta(p,q) distribution are:

$$\mu = \frac{p}{p+q}$$

$$\sigma^2 = \frac{pq}{(p+q)^2(p+q+1)}$$

Solving the first two moments for p and q yields:

$$p = \frac{\mu^2(1-\mu)}{\sigma^2} + \mu$$

$$q = \frac{\mu(1-\mu)^2}{\sigma^2} + (\mu - 1)$$

The mean and variance of transformed residual vector $z_{l,c}$ were substituted into the above two equations to solve for the p and q parameters. The results are as follows:

$$Story\ County\ Corn \sim Beta(5.933, 1.888)$$

$$Story\ County\ Soybeans \sim Beta(3.979, 1.466)$$

[87]The formulas for the minimum and maximum values are based on the first and last order statistics from the Uniform distribution. The estimated minimum and maximum values for the Story county corn yield are 43.87 and 202.59, respectively and the estimated minimum and maximum values for the Story county soybean yield are 27.25 and 58.11, respectively.