A High-Dimensional, Multivariate Copula Approach to Modeling Multivariate Agricultural Price Relationships and Tail Dependencies

Xuan Chi

B.K. Goodwin

Department of Agricultural and Resource Economics

North Carolina State University

xchi@ncsu.edu

barry_goodwin@ncsu.edu

Last updated

August 2012

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2012 AAEA Annual Meeting, Seattle, Washington, August 12-14, 2012

A High-Dimensional, Multivariate Copula Approach to Modeling Multivariate Agricultural Price Relationships and Tail Dependencies

Xuan Chi, Barry Goodwin

I. Introduction

Spatial and temporal relationships among agricultural prices have been an important topic of applied research for many years. Such research is used to investigate the performance of markets and to examine linkages up and down the marketing chain. This research has empirically evaluated price linkages by using correlation and regression models and, later, linear and nonlinear time-series models. The most recent research has recognized the fact that price linkages at different locations or levels of the market may be subject to the influences of adjustment and transactions costs. The result of such costs, which are typically unobservable, is to result in discontinuous and/or nonlinear relationships and patterns of adjustment. The most recent research in this area has recognized the fact that price linkages may be very different during extreme market conditions, such as booms and crashes. This tail behavior has become the focus of a new avenue of empirical research that expresses relationships in terms of tail dependence. Much of this recent research has noted that tail dependence often implies very different economic relationships among those variables that characterize such relationships in more normal states of the market. In addition, extreme shocks in one direction (e.g., busts) may exhibit very different behavior than those occurring in another direction. As a consequence, prices linkages are often observed to have asymmetric tail dependencies.

In our research we use higher dimensional copula models to study regional price linkages in regional soybean and corn markets in multiple spatially distinct regions in North Carolina over the last decade. These markets have been the focus of earlier research (see, for example, Goodwin, Piggott and Sephton (2001)). By estimating multi-dimensional copulas, we are able to formally evaluate the asymmetric tail dependences

among the multiple regional prices of one specific agricultural good. Since transactions costs usually play a significant role in regional price linkages, we expect to see the impacts of such costs in terms of nonlinearities in price linkages. We account for such nonlinearities in a multivariate context by allowing flexibility in the choices of copula functions. We pursue specification testing of alternative copulas and demonstrate approaches to choosing an optimal copula specification by using Cramer von-Mises test statistics. .

Our results provide a means of simultaneously characterizing regional relationships among multiple markets. We demonstrate that these relationships, while statistically significant, are often of a nonlinear nature, corresponding to periods of less than perfect price transmission.

This paper is structured into five parts: the next part is a literature review that provides some background introduction of econometric theory; the third part describes the dataset we use; the fourth part describes models and estimation results; and the final sector presents the conclusion.

II. Literature Review and Background Econometric Theory:

1. Literature Review:

A growing empirical literature has addressed tail dependence among financial time series. Copula models have become popular econometric tools and have drawn extensive attention in evaluations of market linkages. Goodwin et al. (2011) introduced nonlinear time series copula models to evaluate bivariate comparisons of prices among regional commodity markets. However, almost all existing copula-based analyses have been of a pair-wise nature. Even in the rare cases where multidimensional models have been evaluated, the empirical analysis is typically limited to at most five variables. Also, multidimensional models for one single copula are often inflexible. This reflects the fact that higher-dimensional copula models are complex and difficult to implement in many

cases—representing a “curse of dimensionality.” Empirical approaches to evaluating high dimension copula models are a topic of current research.

A very large body of empirical research has examined price relationships within the context of the “law of one price” or “spatial market integration.” Again, most of this research is conducted on pairs of prices rather than considering a wider geographic market. However, it has been observed that collections of pair-wise market relationships also imply multivariate linkages that are fruitful to evaluate. Goodwin (1992) investigated multivariate price linkages among international wheat prices and found that efficiently linked markets imply distinct multivariate relationships.

Research on copulas has recently developed a number of simplified approaches to estimating higher dimensional copula functions. Such approaches include vine copulas and factor copulas. These approaches adopt simplifying restrictions that permit a straightforward evaluation of multivariate copulas. A vine copula is a n-variate parametric copula built by decomposing the multivariate density into a product of bivariate conditional copulas. Aas et al. (2007) developed the method of vine copula estimation, which involves breaking down multivariate copulas into a number of hierarchical bivariate copulas. In addition, Patton (2011) proposed a factor copula approach that enables n-dimension copula estimation in cases where n is large (approaching infinity). Patton provided theoretical proof of the consistency of such estimators and presented empirical applications to stock market returns.

2. Copula Theory:

For an explicit introduction one can refer the recent texts of Nelson (2006) and Joe (1997). By definition, copulas are such functions that join together marginal cdf’s to form multi-dimensional cdf. By separating effects of dependence from effects of margins, copula methods allow one to characterize dependence properties more flexibly.

Let X and Y denote two random variables with joint distribution $F_{X;Y}$ (x; y) and continuous marginal distribution functions $F_X(x)$ and F_Y (y). According to Sklar's (1959) fundamental theorem, there exists a unique decomposition:

$$F_{X;Y}(x; y)=C(F_X(x), F_Y(y))$$

of the joint distribution into its marginal distribution functions and the copula

$$C(u, v) = P(U \leq u, V \leq v), U \equiv F_X(x), V \equiv F_Y(y);$$

defined on [0,1] ×[0, 1] which comprises the information about the underlying dependence structure. Putting a different way, two-dimensional copulas are distribution functions on the unit square with uniform marginals.

A number of different parametric families of copulas are commonly used in analysis of dependence. The two most frequently used parametric copula families are elliptical copulas, which include the Gaussian and Student-t copulas, and Archimedean copulas.

One of the key advantages of copulas lie in their flexibility in measuring tail dependences, which is the dependence between two random variables in the upper-right and lower-left quadrants of their domains (Nelsen 2006). In the case of my study, tail dependence measures how large the association is when one or both the time series has/have large (or small) values.

According to Nelson (2006), the parameter of asymptotic lower tail dependence, noted by λ_L, is the conditional probability in the limit that one series takes a very low value, given that the other also takes a very low value. Similarly, the parameter of asymptotic upper tail dependence, noted by λ_U, is the conditional probability in the limit that one series takes a very high value, given that the other also takes a very high value. The asymptotic tail dependence parameters for copula function are shown as following (Nelsen 2006):

$$\lambda_L = \lim_{t \to 0+} \left(\frac{C(t, t)}{t} \right)$$

$$\lambda_U = 2 - \lim_{t \to 1-} \left(\frac{1 - C(t,t)}{1 - t} \right)$$

In the case of Gaussian copula and Student-t copula, the copula functions are symmetric, which implies that the asymptotic upper and lower tail dependences are identical. Note that most of the copula families have asymmetric upper and lower tail dependences.

Conventional copulas usually consider only bivariate distributions of two data series. The definition of vine copulas was introduced by Joe (1997). A vine copula is a n-variate parametric copula built by decomposing the multivariate density into a product of bivariate copulas.

For example, a trivariate (3 dimensional) copulas can be written as:

$$C_{123} = C_{12}\big(F_1(x_1);\ F_2(x_2)\big) \times C_{13}\big(F_1(x_1);\ F_2(x_2)\big)$$

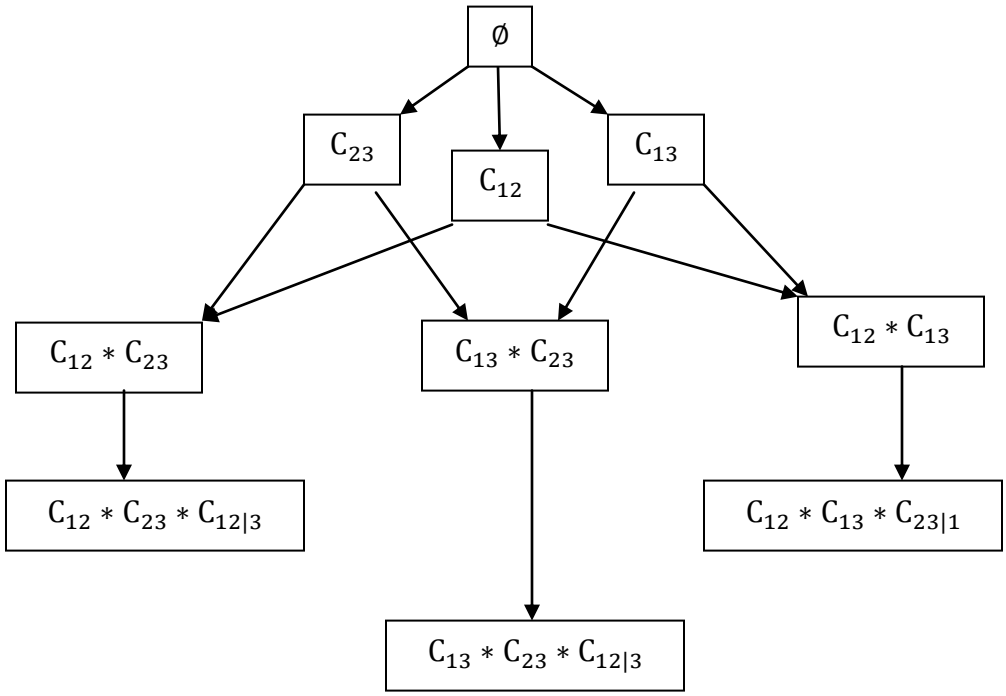
$$\times\ C_{23|1}(\partial_1 C_{12}\big(F_1(x_1);\ F_2(x_2)\big); \partial_1 C_{13}\big(F_1(x_1);\ F_2(x_2)\big))$$

Or in short:

$$C_{123} = C_{12} \times C_{13} \times C_{23|1}$$

We can see that in this decomposition multivariate copulas are broken down into several hierarchical bivariate copulas, which is a practical way to represent high dimensional copula problems. Please see the following picture for a whole idea of vine copulas in this example.

Figure 1: (see Maugis, Guegan 2010)



Aas et al. (2007) provided detailed discussions about the development and application of vine copulas. They used the pair-copula decomposition of a general multivariate distribution and propose a method to perform inference.

III. Data

Daily corn prices data for spatially distinct North Carolina markets were collected and maintained by the North Carolina Department of Agriculture[1]. The data covers prices in nearly all the important corn markets in North Carolina, ranging from 1976 to 2011. We selected five cities among them that have no more than 5% missing values: Candor, Cofield, Roaring River, Statesville and Pantego. Also we constrained the time range to

[1] The price data are compiled by Anton Bekkerman and Nick Piggott.

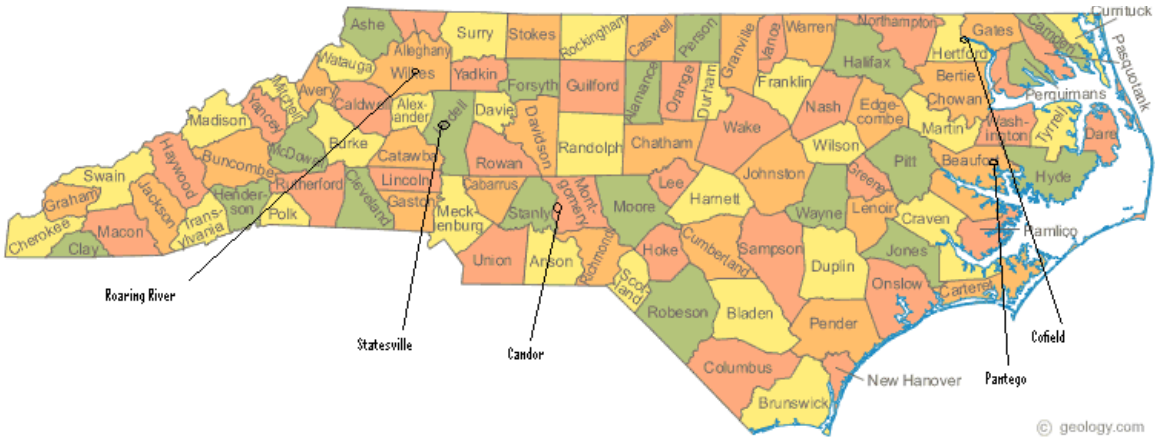
the decade of 2001-2010. The following table is the descriptive statistics of the corn prices dataset.

Table 1: Some Descriptive Statistics:

	City	Corn Prices($)
N		2597
Mean	Candor	3.29
	Cofield	3.07
	Roaring River	3.22
	Statesville	3.16
	Pantego	2.91
Variance	Candor	1.57
	Cofield	1.32
	Roaring River	1.33
	Statesville	1.21
	Pantego	0.86
Range	Candor	[1.84, 8.62]
	Cofield	[1.58, 8.47]
	Roaring River	[1.75, 8.62]
	Statesville	[1.8, 8.43]
	Pantego	[1.79, 6.2]

The five cities on the map of North Carolina are shown below as Figure 2. We can see that the five cities are displayed randomly on the map. Three cities lie on the west of North Carolina while the other two lie on the east.

Figure 2: Map of North Carolina:



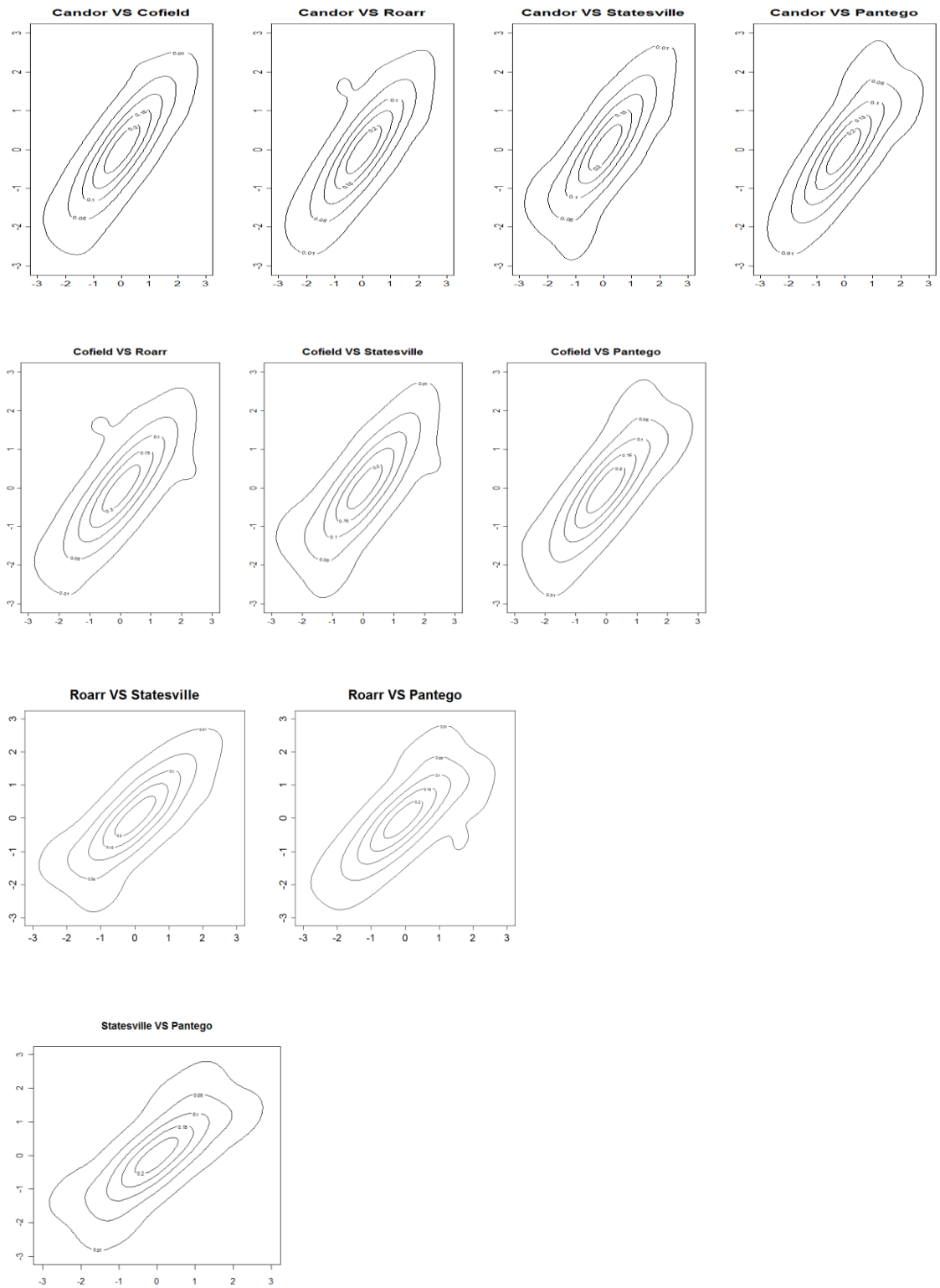
IV. Model

(1) Contour Set Plots

A contour set plot is a useful tool to illustrate intuitive relationship among multiple data series. The following plots are the bivariate contour plots of corn prices among the five cities.

By simple calculation we know there are 10 pairs of contour sets among the five cities. The four plots in the first row represent the combinations of: Candor VS Cofield, Candor VS Roaring River, Candor VS Statesville, and Candor VS Pantego. The three plots in the second row are the combinations of Cofield VS Roaring River, Cofield VS Statesville, and Cofield VS Pantego. The third row is similarly illustrated as the two combinations of Roaring River VS Statesville and Roaring River VS Pantego. The last row is the last pair of Statesville VS Pantego.

Figure 3: Empirical contour sets for 10 pairs of corn prices



We can get a first and general impression that there exist tail dependences among all the pairs of corn prices. Independence tests also support the same conclusion. Therefore copula models can be a good fit for our dataset.

(2) Bivariate Copula Estimation:

In order to estimate higher dimensional copula functions, we firstly estimated bivariate copula functions among the 10 pairs of corn prices in the five cities. The estimation results are as the following table, with estimated parameters in the lower triangle cells. Here the two parameters in the parentheses are the copula parameters. The copula models are selected based on the rankings of AIC or BIC functions.

Table 2: Bivariate copula estimation for the 10 pairs of corn prices

	Candor	**Cofield**	**Roarr**	**Statesville**	**Pantego**
Candor	-	Frank	T	Frank	Frank
Cofield	(36.955, 0)	-	T	Frank	Frank
Roarr	(0.9887, 2)	(0.9819, 2)	-	Frank	Frank
Statesville	(23.406, 0)	(22.629, 0)	(21.252, 0)	-	Frank
Pantego	(32.669, 0)	(37.658, 0)	(22.264, 0)	(21.349, 0)	-

We can see that most of the bivariate copulas are Frank copulas, while only two of them are T copulas. These results are also consistent with the Vuong Clarke Tests. Most of the Vuong Clarke Tests results lead to T or Frank copula models specification. Therefore, we know that most of the copula models for corn prices would be symmetric copulas.

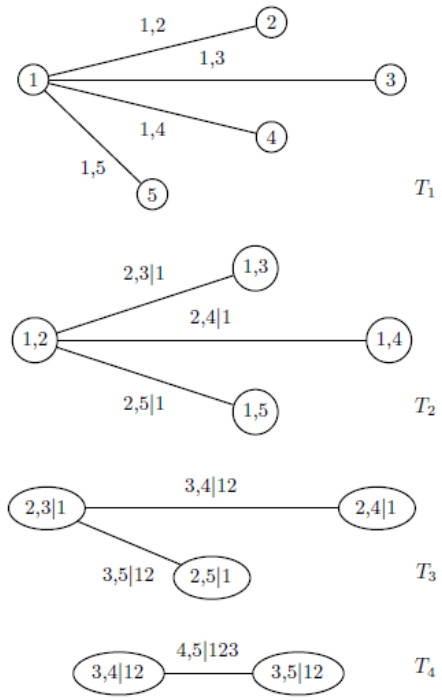
Although the bivariate copulas are useful in capturing pair-wise relationships, we still expect the vine copula models to provide us a more flexible and parsimonious multidimensional structure.

(3) C- and D-vine copula estimation of all five prices:

There are two types of commonly used vine structures, canonical vine copulas (C-vine) and D-vine copulas. Based on Vuong Clarke Tests for the vine copulas, we chose C-vines instead of D-vines. Actually, the Vuong Clarke tests for them do not show significant differences between them, so we attached the D-vine estimation results in the Appendix.

For a five-dimensional C-vine copula, the tree structure should be like the following table. Note that the (1; 3|2) means the copula model between variable 1 and variable 3 conditional on variable 2. More specifically, tree 1 is the bivariate copula for the four price couples, and tree 2 is the bivariate copula for the three price couples conditional on the third one, and tree 3 and 4 follow with the same principle.

Figure 4: Tree structure of C-vine copula (see CDVine R package)



In other words, the final results of our C-vine copula model would a tree structure with four layers of conditional copulas shown in Table 3.

Table 3: Tree structure of C-vine copula

(1; 2); (1; 3); (1; 4); (1; 5)	Tree1
(2; 3|1); (2; 4|1); (2; 5|1)	Tree2
(3; 4|1; 2); (3; 5|1, 2)	Tree3
(4; 5|1; 2; 3)	Tree4

Based on the method suggested by Czado (2011), we select Candor as the first root node of the tree by maximizing the bivariate dependencies. Sequentially, we can determine the following four nodes. The tree nodes are orderd as: 1) Candor; 2) Cofield;

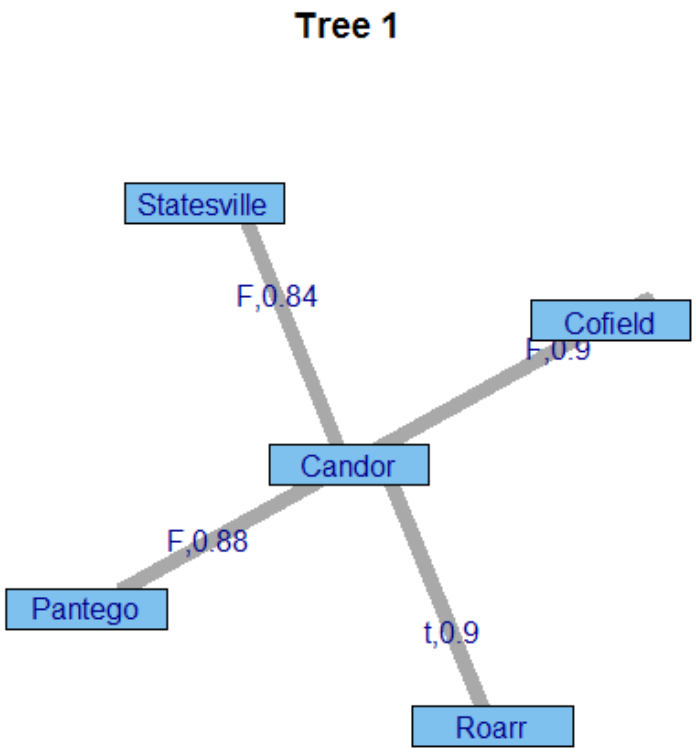
3) Statesville; 4) Roaring River; 5) Pantego. We will use number 1 to 5 to denote the five corn prices in all the following results report.

Table 4: vine copula estimation for Tree 1

Tree 1	(1; 2)	(1; 3)	(1; 4)	(1; 5)
Family	F*	F*	T*	F*
Par1	36.9546	28.1401	0.9615	32.6694
(se)	(0.6671)	(0.5425)	(0.0017)	(0.5939)
Par2	0	0	2	0
(se)	(0.0000)	(0.0000)	(0.1095)	(0.0000)

*F=Frank copula; **T= Student's T copula

Figure 5: Tree plot for Tree 1



The above estimation results and tree plot show us the first layer of our vine copula model, which we called Tree 1. We can see the five combinations of bivariate copulas are estimated, and their Kendall's Taus are also shown on the picture respectively.

Next we will see the estimation results for Tree 2 as in Table 5 and Figure 6:

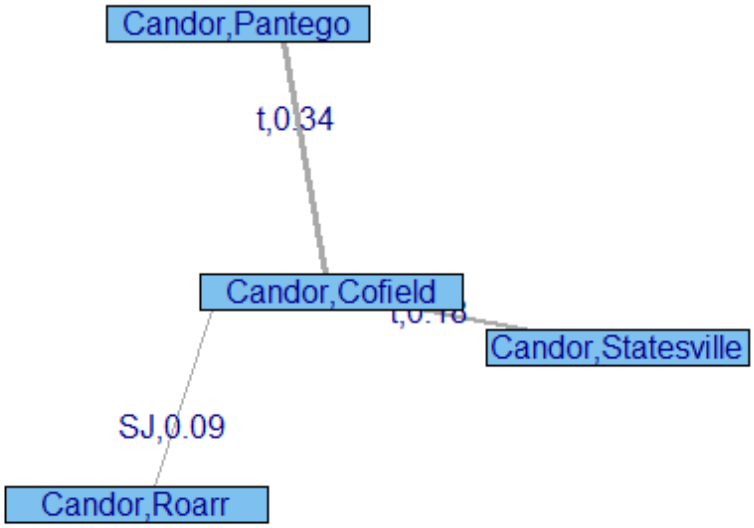
Table 5: vine copula estimation for tree 2:

Tree 2	(2; 3|1)	(2; 4|1)	(2; 5|1)
Family	T**	SJ##	T**
Par1	0.3316	1.0422	0.5103
(se)	(0.0240)	(0.00853)	(0.0157)
Par2	12.1356	0	8.9255
(se)	(1.1522)	(0.0000)	(0.9698)

**T= Student's T copula; ##SJ=Survival Joe copula, or Joe copula rotated 180 degrees.

Figure 6: Tree plot for Tree 2

Tree 2



In Tree 2 we can see three conditional copula models. Cofield VS Pantego given Candor (2; 5|1) is estimated as a Student's t copula, Cofield VS Roaring River given Candor (2; 4|1) is estimated as Survival Joe copula, and Cofield VS Statesville given Candor (2; 3|1) is estimated as Student's t copula again. Their Kendall's Taus are also shown in the picture.

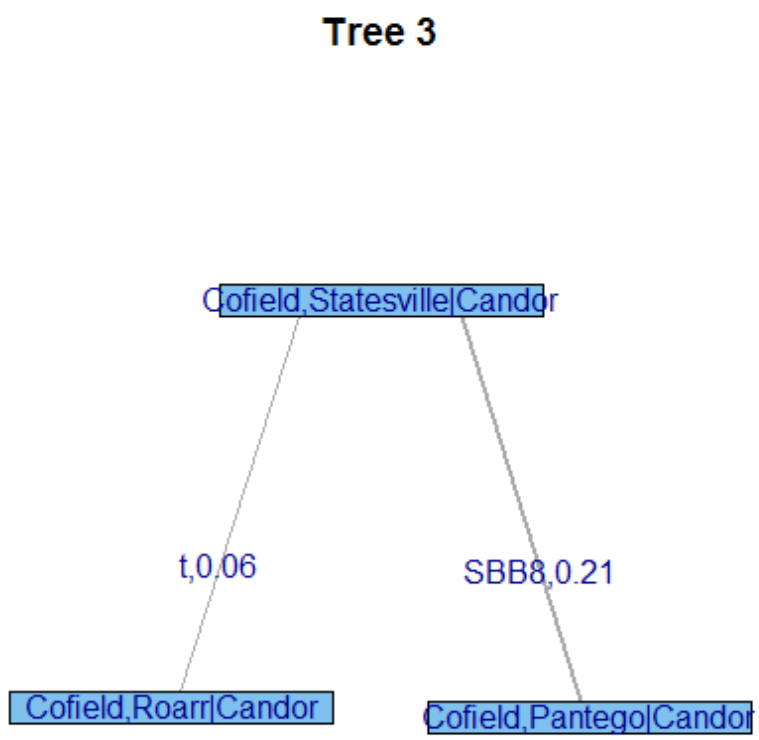
Table 6: vine copula estimation for Tree 3

Tree 3	(3; 4|1; 2)	(3; 5|1, 2)
Family	SBB8#	T**
Par1	0.09772	6
(se)	(0.0200)	(2.2820)

Par2	22.2850	0.3134
(se)	(2.6058)	(0.1127)

**T= Student's T copula; #SBB8=Survival Joe-Frank copula, or Joe-Frank copula rotated 180 degrees.

Figure 7: Tree Plot for Tree 3



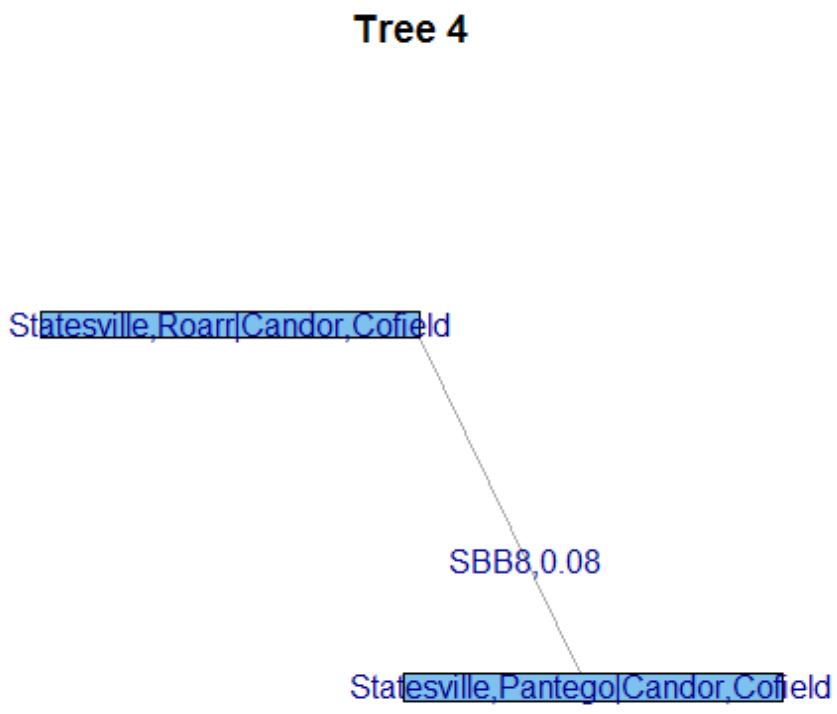
Similarly, we can judge from the above results that two conditional copulas are estimated. Roaring River VS Statesville given Cofield and Candor (3; 4|1; 2) is estimated as Survival Joe-Frank Copula, while Roaring River VS Pantego given Cofield and Candor (3; 5|1; 2) is estimated as T copula.

Table 7: vine copula estimation for Tree 4:

Tree 4	(4; 5|1; 2; 3)
Family	SBB8##
Par1	1.4343
(se)	(0.1517)
Par2	0.7626
(se)	(0.0889)

##SBB8=Survival Joe-Frank copula, or Joe-Frank copula rotated 180 degrees.

Figure 8: Tree Plot for Tree 4



The last one is the most "complicated" conditional copula, Statesville VS Pantego given Candor, Cofield and Roaring River (4; 5|1: 2; 3), which is estimated as a Survival Joe-Frank copula model.

The entire conditional copula models above shed lights to new copula families such as Joe-Frank and Survival Joe-Frank copulas, along with their estimated parameters. For more detailed description of each copula family, please refer to Joe (1997).

V. Conclusion

In this paper we modeled the corn prices in five main markets in North Carolina with higher dimensional copula specification, which can give us a detailed description of the dependencies among the five prices and the tree plot showed us an intuitive correlation among them. As the results show, bivariate copula models tend to have symmetric structures such as T and Frank copulas, while the C-vine copula estimation give us more copula models specification(such as BB8, SBB8) in a hierarchal structure.

References

B.K. Goodwin and Piggott (2001), *Dynamic linkages among real interest rates in international capital markets*, American Journal of Agricultural Economics

Goodwin et al. (2011), *Copula-Based Nonlinear Models of Spatial Market Linkages*, 2011 AAEA Annual Summer Meetings.

Goodwin (1992), *The Capitalization of Wheat Subsidies into Agricultural Land Values*,

Canadian Journal of Agricultural Economics

Aas et al. (2007) *Pair-copula constructions of multiple dependence*, Insurance: Mathematics and Economics

Czado (2011) *Maximum likelihood estimation of mixed C-vines with application to exchange rates*, Statistical Modelling

Patton (2011), *Copula Methods for Forecasting Multivariate Time Series*

Joe (1997) *Multivariate models and dependence concepts*, Chapman & Hall.

Nelson (2006) *An Introduction to Copulas*, Springer Series in Statistics

Brechmann, Czado (2012) *Financial risk management with high-dimensional vine copulas*, Working Paper

Maugis, Guegan (2010) *Dealing with vine issues*, 3rd financial risks international forum

Appendix: D-vine copula estimation results:

Table 1: Tree structure of D-vine copula

(1; 2); (2; 3); (3; 4); (4; 5)	Tree1
(1; 3|2); (2; 4|3); (3; 5|4)	Tree2
(1; 4|2; 3); (2; 5|3, 4)	Tree3
(1; 5| 2; 3; 4)	Tree4

Based on the tree structure above, we estimated their copula families and copula model parameters. Please see following tables and figures as the results.

Table 2: vine copula estimation for Tree 1

Tree 1	(1; 2)	(2; 3)	(3; 4)	(4; 5)
Family	F*	T**	F*	F*
Par1	36.955	0.9819	21.2518	21.3486
(se)	(0.6671)	(0.000814)	(0.3948)	(0.3892)
Par2	0	2	0	0
(se)	(0.0000)	(0.1296)	(0.0000)	(0.0000)

*F=Frank copula; **T= Student's T copula; # BB8= Joe-Frank copula; ##SBB8=Survival Joe-Frank copula, or Joe-Frank copula rotated 180 degrees.

Figure 1: Tree plot for Tree 1

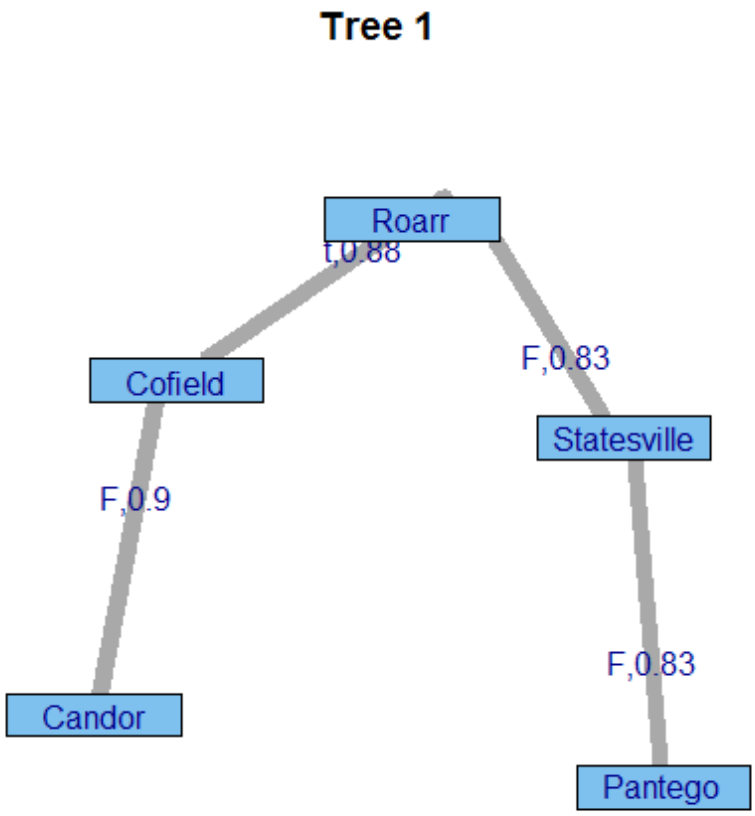


Table 3: vine copula estimation for tree 2:

Tree 2	(1; 3|2)	(2; 4|3)	(3; 5|4)
Family	SBB8##	T**	SBB8##
Par1	6.000	0.28852	3.38535
(se)	(0.46535)	(0.01792)	(0.20237)
Par2	0.7105	15.33543	0.90585
(se)	(0.03162)	(2.0449)	(0.02105)

*F=Frank copula; **T= Student's T copula; # BB8= Joe-Frank copula; ##SBB8=Survival Joe-Frank copula, or Joe-Frank copula rotated 180 degrees.

Figure 2: Tree plot for Tree 2

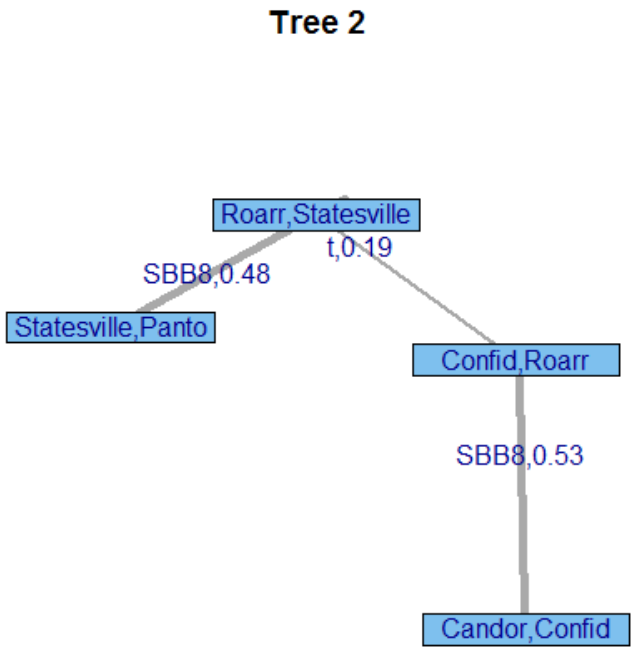


Table 4: vine copula estimation for Tree 3

Tree 3	(1; 4|2; 3)	(2; 5|3, 4)
Family	BB8#	T**
Par1	1.9560	0.39791
(se)	(0.3877)	(0.0164)
Par2	0.73415	10.7368
(se)	(0.1239)	(1.2029)

*F=Frank copula; **T= Student's T copula; # BB8= Joe-Frank copula;

##SBB8=Survival Joe-Frank copula, or Joe-Frank copula rotated 180 degrees.

Figure 3: Tree Plot for Tree 3

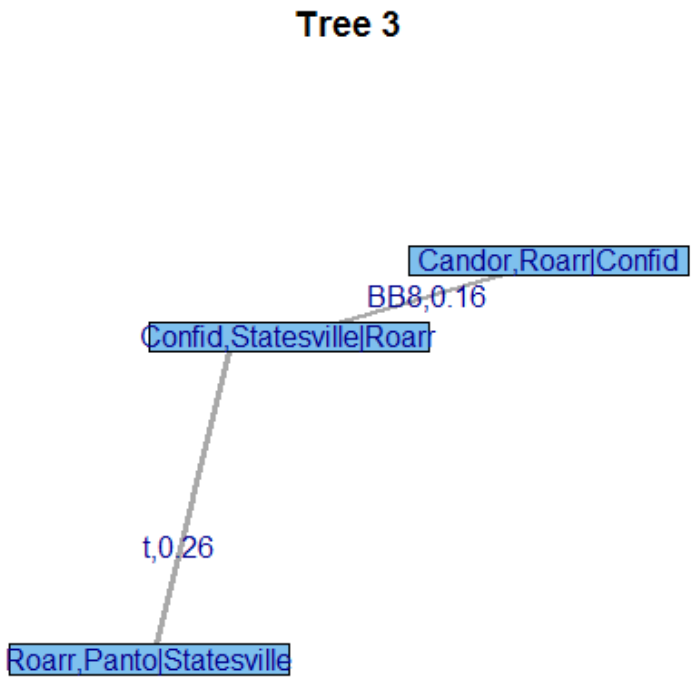


Table 5: vine copula estimation for Tree 4:

Tree 4	(1; 5| 2; 3; 4)
Family	SBB8##
Par1	1.99765
(se)	(0.2702)
Par2	0.7190
(se)	(0.0771)

*F=Frank copula; **T= Student's T copula; # BB8= Joe-Frank copula; ##SBB8=Survival Joe-Frank copula, or Joe-Frank copula rotated 180 degrees.

Figure 4: Tree Plot for Tree 4