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

June 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 results 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 than those that characterize relationships in normal states of the market. In addition, extreme in one direction (e.g., busts) may exhibit very different behavior than in another direction. As a consequence, prices linkages are often observed to have asymmetric tail dependencies.

In our research we use the higher dimensional copula models to study regional price linkages of regional soybean and corn markets in multiple regions in North Carolina over the last decade. These markets have been the focus of earlier research (see, for example, Goodwin and Piggott (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 among, 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 first one is introduction; the second part is a literature review and some background introduction of econometric theory; the third part described the dataset we used; the fourth part described models and estimation results; finally it comes to 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 agricultural 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. 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, by Aas et al.(2007). 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. About Copula Theory:

For an explicit introduction one can refer to two books: An Introduction to Copulas by Nelson (2006), and Multivariate Models and Dependence Concepts by Joe (1997).

In 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 people 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 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 copula methods is its 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 are functions only about bivariate distributions of two data series. Then the definition of vine copulas was introduced (Joe 1997). A vine copula is a n-variate parametric copula built by decomposing the multivariate density in a product of bivariate copulas.

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

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

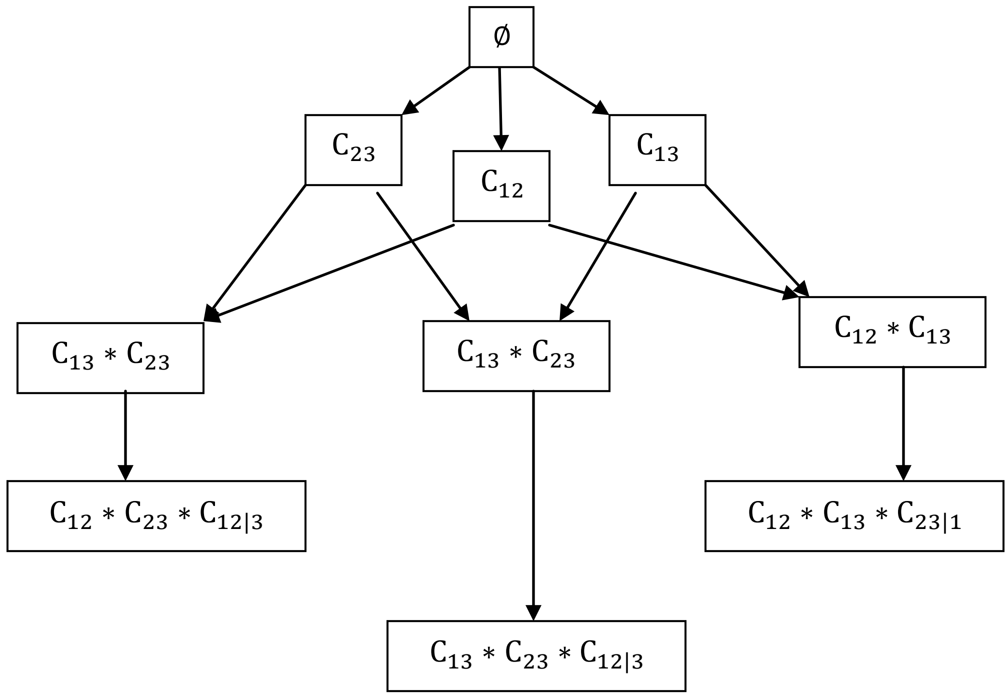
$$\times\, C_{23|1}(\partial_1 C_{12}\big(F_1(x_1);\ F_2(x_2)\big); \partial_1 C_{13}\big(F_1(x_1);\ F_2(x_2)\big))$$

Or in short:

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

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

Picture 1:



Aas et al(2007) provided discussions about applications 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 were collected and maintained by Anton Bekkerman and Nick Piggott. 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, Roarring River, Statesville and Panto. Also we constrained the time range to the decade of 2001-2010. The following table is the descriptive statistics of the corn prices dataset.

Table 1: Some Descriptive Statistics:

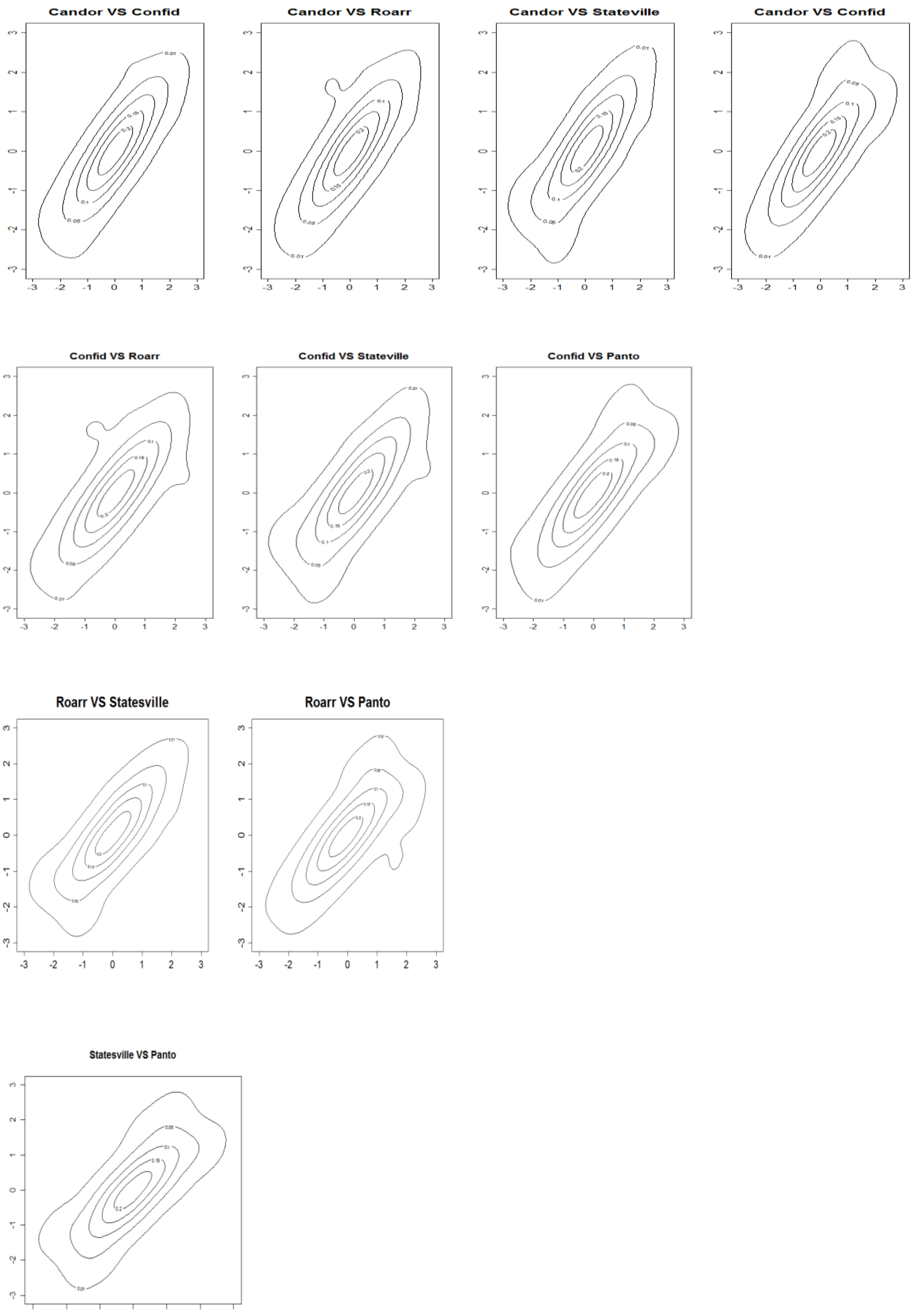
	City	Corn Prices (dollar)
N		2597
Mean	Candor	3.29
	Cofield	3.07
	Roarr	3.22
	Statesv	3.16
	Panto	2.91
Variance	Candor	1.57
	Cofield	1.32
	Roarr	1.33
	Statesv	1.21
	Panto	0.86
Range	Candor	[1.84,8.62]
	Cofield	[1.58,8.47]
	Roarr	[1.75,8.62]
	Statesv	[1.8,8.43]
	Panto	[1.79,6.2]

IV. Model

(1) Contour Set Plots

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.

Figure 2:


Candor VS Confid
Candor VS Roarr
Candor VS Stateville
Candor VS Confid
Confid VS Roarr
Confid VS Stateville
Confid VS Panto
Roarr VS Statesville
Roarr VS Panto
Statesville VS Panto

We can get a first and general impression that there exist tail dependences among all the pairs of corn prices. Independence tests also supported the same conclusion. Therefore copula models can be 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 corn prices of the five cities. The estimation results are as following table, with estimated parameters in the lower triangle cells.

Table 2:

	Candor	**Confid**	**Roarr**	**Statesville**	**Panto**
Candor	-	Frank	T	Frank	Frank
Confid	(36.814, 0)	-	T	Frank	Frank
Roarr	(0.9887, 2)	(0.9818, 2)	-	Frank	Frank
Statesville	(23.406, 0)	(22.569, 0)	(21.252, 0)	-	Frank
Panto	(32.669, 0)	(37.515, 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 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.

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

We chose D-vine copula estimation instead of canonical vine copulas(C-vine) because it has a symmetric structure, while C-vine requires a hierarchical node selection process. Since the corn prices in different locations should express a similar and symmetric relationship, we are confident that the D-vine copulas are better fit for our model. This can also be approved by the Vuong Clarke Tests for vine copulas.

For a five-dimensional D-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.

Table 3: Tree structure

(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|1; 2; 3; 4)	Tree4

Based on the tree structure above, we estimated their copula families and copula model parameters. Please see table 3 as the results.

Table 3: vine copula estimation

	(1; 2)	(2; 3)	(3; 4)	(4; 5)	(1; 3|2)	(2; 4|3)	(3; 5|4)	(1; 4|2; 3)	(2; 5|3, 4)	(1; 5|1; 2; 3; 4)
Family	F*	T	F	F	SBB8**	T	SBB8	BB8#	T	SBB8
Par1	36.814	0.982	21.252	21.349	6.00	0.287	3.385	1.944	0.398	2.003
Par2	0	2	0	0	0.711	15.28	0.906	0.739	10.74	0.718

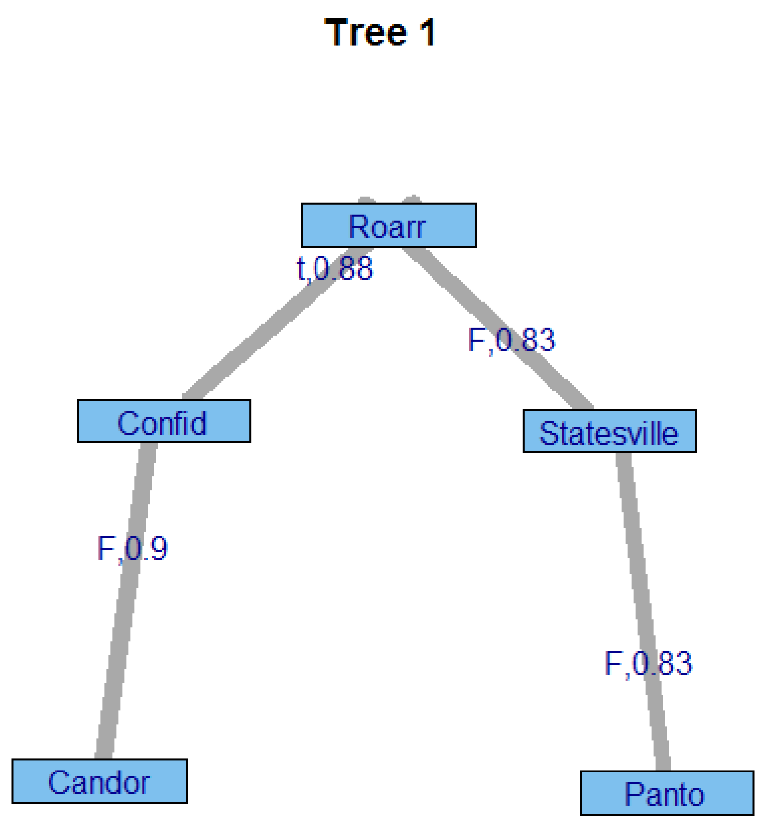
*F=Frank copula; **SBB8=Survival Joe-Frank copula, or Joe-Frank copula rotated 180 degrees; # BB8= Joe-Frank copula

Now the conditional copula estimation gives us new families such as Joe-Frank and Survival Joe-Frank copulas, and their parameters are also given. For more detailed description of each copula families, please refer to Joe (1997).

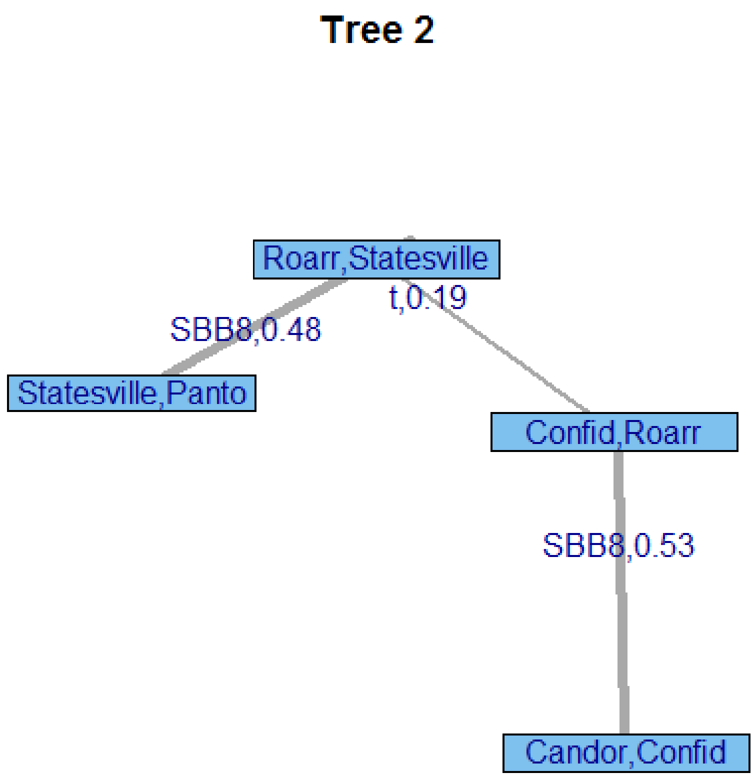
(4) Tree Plots:

With the estimation above, now we are able to plot the tree structure of our vine copula model. Please see the following pictures as the trees in different levels.

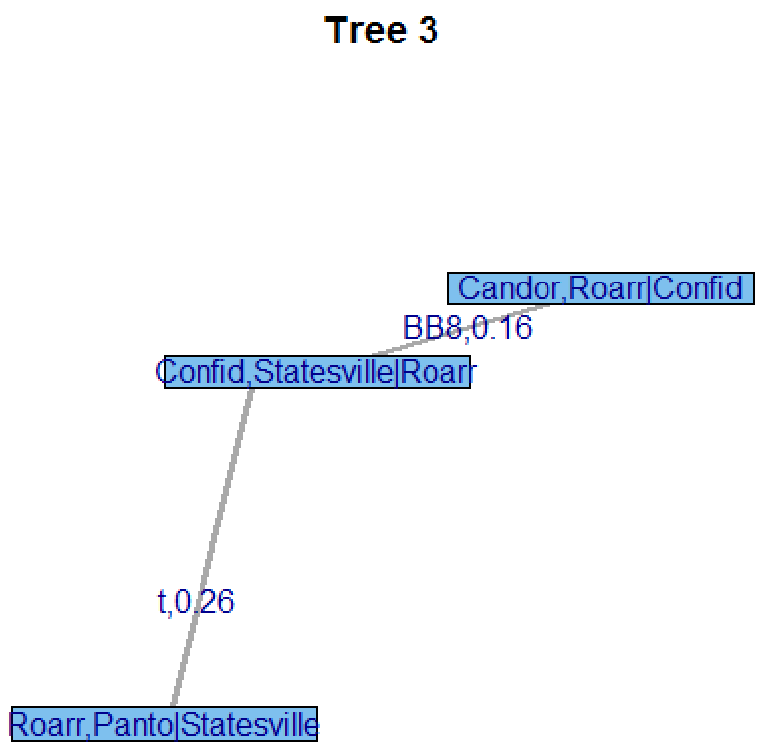
Picture 3:


Tree 1
Roarr
t,0.88
F,0.83
Confid
Statesville
F,0.9
F,0.83
Candor
Panto

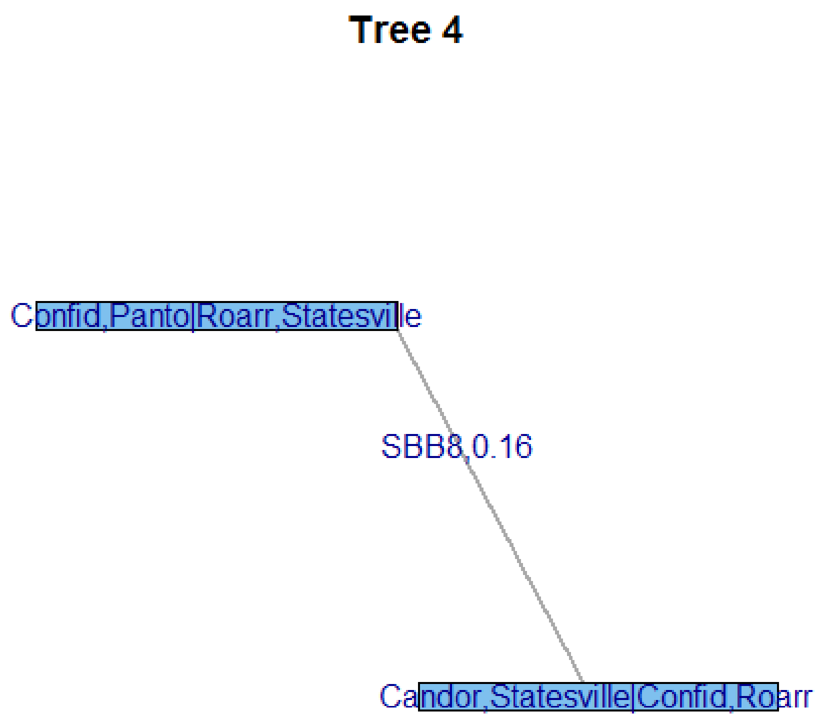
Picture 4:


Tree 2
Roarr,Statesville
SBB8,0.48
t,0.19
Statesville,Panto
Confid,Roarr
SBB8,0.53
Candor,Confid

Picture 5:


Tree 3
Candor,Roarr|Confid
BB8,0.16
Confid,Statesville|Roarr
t,0.26
Roarr,Panto|Statesville

Picture 6:



These four tree plots illustrated the D-vine copula models' structure among the five price series, as explained in Table 3.

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 D-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

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