

A Copulas-based Approach to Modeling Dependence in Decision Trees

Tianyang Wang¹ and James S. Dyer²

¹College of Business, Colorado State University, Fort Collins, Colorado, 80523

²McCombs School of Business, The University of Texas at Austin, Austin, Texas 78712

{¹tianyang.wang@business.colostate.edu, ²jim.dyer@mcombs.utexas.edu}

Abstract:

This paper presents a general framework based on copulas for modeling dependent multivariate uncertainties through the use of a decision tree. The proposed dependent decision tree model allows multiple dependent uncertainties with arbitrary marginal distributions to be represented in a decision tree with a sequence of conditional probability distributions. This general framework could be naturally applied in decision analysis and real options valuations, as well as in more general applications of dependent probability trees. While this approach to modeling dependencies can be based on several popular copula families as we illustrate, we focus on the use of the normal copula and present an efficient computational method for multivariate decision and risk analysis that can be standardized for convenient application.

Key words: correlation; copulas; multivariate decision and risk analysis

1. Introduction

Modeling uncertainties is a crucial part of decision and risk analysis. In reality, multiple sources of uncertainty commonly exist, and they are likely to be dependent. With the advantage of a visual interface and natural backward dynamics for early exercise options, the decision tree has become one of the fundamental tools for modeling uncertainties; however, one weakness in its practical use is that it is computationally difficult to incorporate dependencies among different uncertainties in the decision tree. Many times, these dependencies are neglected in modeling in order to simplify the analysis as discussed in Abbas(2006) and Bickel and Smith(2006). For situations where such negligence may cause significant errors, the incorporation of these dependencies into the decision and risk probability models becomes important (Smith, et al., 1992).

The most natural approach to modeling multivariate uncertainties in a decision tree is to specify their conditional distributions given the discrete outcomes of the preceding uncertainties. Despite its simple logic, this approach is limited in practical applications because conditional distributions are not easy to derive analytically, especially when the marginal distributions are from different parametric families. Additionally, the task of specifying all of the necessary conditional distributions grows combinatorially with the number of variables; therefore, this approach is information-intensive and impractical for subjective risk assessment when empirical knowledge is limited (Ghosh and Henderson, 2002).

A practical alternative is to specify only the marginal distributions of the uncertainties and to describe the dependencies among them using measures of correlation or of tail dependence, and then to calculate the corresponding joint distribution using an appropriately chosen copula. A copula function links the univariate marginal distributions with their multivariate joint distribution, and the choice of a specific copula function allows flexibility in modeling different dependence relationships. There are many functional forms of copulas that are grouped into families, and among the most commonly used are those from the elliptical and Archimedean families (Nelsen, 1999; Trivedi and Zimmer, 2005). The popular normal copula and the t-copula are members of the elliptical family of copulas.

The most commonly used measure of dependence among marginal distributions is correlation, and typically this dependence is measured with the familiar Pearson product moment correlation coefficient. While the Pearson correlation measure is convenient to use and appropriate in many applications, it is an estimate of the linear relationship between two marginal

distributions, and may therefore be inaccurate if the correlation relationship is non-linear. In such cases, Spearman's rank correlation coefficient or Kendall's tau correlation measure may be used, since they are based on rank data and therefore do not assume a linear relationship between the two marginal distributions. Copula functions exist that can be used with all three of these correlation measures, and therefore provide flexibility in modeling statistical relationships. Marginal distributions and correlation measures are all that are needed by the family of elliptical copulas to determine the corresponding joint distributions of the uncertain variables.

However, there are other forms of dependence between marginal distributions that are not captured by correlations, and that can be important in some modeling contexts. One such consideration in modeling probabilistic dependency is the existence of tail dependency, where there may be a strong probabilistic relationship between extreme values of two marginal distributions. For example, there may be a stronger dependence between large losses than between large gains, as we have observed in stock price movements when the stock market experiences extreme losses (Malevergne and Sornette, 2005), or damage losses among different lines of insurance when catastrophes occur (Kousky and Cooke, 2009). Tail dependencies have also been identified in oil and gas applications (Al-Harthy et al., 2007). When such situations arise, Archimedean copulas that capture asymmetrical tail dependence may be used as more appropriate modeling tools.

Copulas have been applied widely in risk analysis (e.g., Embrechts, et al., 1999), insurance (e.g., Kousky and Cooke 2009), and finance (e.g., Cherubini et al., 2004; Biller, 2009), where the emphasis has been on the use of Monte Carlo simulation models to implement them to generate multivariate joint distributions. The use of copula-based Monte Carlo simulation has also been suggested to support decision analysis applications in the oil and gas industry by Accioly and Chiyshi (2004), Armstrong, et al. (2004), and by Al-Harthy, et al. (2007). In contrast, we will illustrate how copulas can be used to construct probability trees composed of discrete, conditional probability nodes that can be used as the building blocks for decision trees. However, the construction of these discrete, conditional probability trees is a general process, and they could be used in other applications in finance or risk analysis as well.

The application of copulas to problems in decision analysis involving discrete probability trees has been limited. Clemen and Reilly (1999; hereafter, C&R) developed a feasible method to model dependence in the decision tree using just marginal distributions and correlation information. They showed how to construct a normal copulas-based joint distribution model to

calculate the conditional densities and to use the discrete approximation of these conditional densities to create a multivariate decision tree.

A disadvantage of the C&R approach is that the computational requirements associated with these calculations are often challenging and may require customized coding for each application. In this paper, we propose a mathematically equivalent yet computationally more efficient alternative to the C&R approach that shares the benefits of this approach and reduces its practical limitations. When the dependency structure is modeled by the normal copulas, the dependent decision tree approach extends NORTA (NORmal To Anything) (Cario and Nelson 1997), a popular approach developed for the support of high dimensional Monte Carlo simulation models, to modeling dependence in decision trees (see details in the Appendix 1).

The remainder of this paper is organized as follows: Section 2 discusses modeling multivariate distribution with copulas. Section 3 presents how to construct the copula based dependent decision trees. Section 4 discusses the computational advantage of the dependent decision tree approach. An airline example is presented to demonstrate the dependent decision tree approach and to provide a comparison with the C&R approach. Section 5 presents the consolidated dependent decision tree, an extension of the dependent decision tree approach for multivariate real options valuation. An oil industry example is used to demonstrate the consolidated dependent decision tree extension. Section 6 provides a summary of the work and discusses its limitations.

2. Discussion of Copulas

A copula allows a joint distribution of random variables to be expressed as a function of the marginal distributions. For simplicity, we consider the bivariate case. For any real valued random variables X_1 and X_2 with marginal cumulative distribution functions (CDF) $F_1(X_1)$ and $F_2(X_2)$, respectively, Sklar (1959) showed that there always exists a function C such that

$$F_{12}(X_1, X_2) = C(F_1(X_1), F_2(X_2)) \quad (2.1)$$

The function C is called a copula. It fully captures the dependence structure among the random variables, and is independent from the choice of the marginal distributions. Therefore, the joint distribution can be constructed from two independent components: the copula and the marginal distributions. The marginal distributions need not be from the same distribution family, and they are “coupled together” using a copula function. While the individual uncertainties involved in applications of copulas in finance and insurance often have marginal distributions from the same family, this may not be true in practical applications of decision and risk analysis.

The copula function C is itself a distribution function for uniform random variables since the marginal CDFs are standard uniform distributions. Suppose that continuous random variables X_1 and X_2 have a joint distribution function $F_{12}(X_1, X_2)$ defined by (2.1), and let $U_1 = F_1(X_1)$ and $U_2 = F_2(X_2)$; the joint CDF can be given by the copula

$$C(u_1, u_2) = F_{12}(F_1^{-1}(u_1), F_2^{-1}(u_2)) \quad (2.2)$$

where F^{-1} denotes the inverse cumulative distribution function, and $u_i, i = 1, 2$ is an uniform variable defined over $[0, 1]$.

This property suggests an effective algorithm for constructing decision trees by focusing on the dependence structure defined by copulas with unified standard uniform variables. Because the copula contains all information regarding the dependence structure, we can first construct a transient probability tree independent of the choice of marginal distributions for the underlying copula, and then combine it with the information about the marginal distributions to produce the desired conditional distributions in the constructed dependent decision tree.

The transient probability tree structure for the underlying copula models the standard uniform pair u_1 and u_2 separately. To generate the dependent uniform random variable u_2 given each realization of $u_1 = \Pr(X_1 \leq x_1) = \alpha_1$, we will calculate the conditional distribution of X_2 given $X_1 = x_1$ from the partial derivative of the copula

$$\Pr(X_2 \leq x_2 | X_1 = x_1) = \frac{\partial C}{\partial F_1(x_1)}(F_1(x_1), F_2(x_2)) = \frac{\partial C}{\partial u_1}(u_1, u_2) \quad \text{Darsow et al. (1992)}$$

For each given conditional percentile $\Pr(X_2 \leq x_2 | X_1 = x_1) = \alpha_2$, the dependent uniform variable $u_2 = \Pr(X_2 \leq x_2)$ is the inverse function of the realization of $u_1 = \alpha_1$ and the choice of the percentile α_2 for the conditional distribution $X_2 \leq x_2 | X_1 = x_1$. Let $\frac{\partial C}{\partial u_1}(u_1, u_2) = c_{u_1}(u_2) = \alpha_2$, then $u_2 = c_{u_1}^{-1}(\alpha_2)$. Therefore, the transient probability tree can be constructed with the unconditional percentile α_1 and the choice of the conditional percentile α_2 . The transient tree structure is then transformed into the desired dependent decision tree as we will illustrate in the next section. The formulae for the copula, the conditional distribution via the copula, and for u_2 provide the information necessary to implement a specific copula for use in developing a dependent decision tree.

The elliptical copulas and Archimedean copulas are among the most popular copula families in applied modeling and will be the focus of this research. The elliptical copulas and the Archimedean copulas differ significantly in modeling the tail dependency of distributions.

Tail dependency measures the probability that extreme events happen jointly. Upper tail dependence exists when there is a probability that positive extreme events happen jointly. Upper tail dependence is defined as:

$$\lambda_{upper} = \lim_{u \rightarrow 1} \Pr(X_2 \geq F_{X_2}^{-1}(u) | X_1 \geq F_{X_1}^{-1}(u)) \quad (2.3)$$

Lower tail dependence is defined symmetrically. The tail dependency measure depends only on the copula and not on the marginal distributions.

Elliptical copulas are restricted to symmetrical tail dependencies. They are simply the copulas of elliptical distributions, and provide useful examples of multivariate distributions because they share many of the tractable properties of the elliptical distributions. The normal copula and the t-copula belong to the elliptical copulas family. The key formulae for the normal and t-copulas are summarized in Table 1 (Cherubini et al., 2004).

Table 1 Key Formulae for Elliptical Copulas

Family	Copula Description	Partial derivative of the copula	u_2
normal copula	$C_N(F_1(X_1), F_2(X_2)) = \Phi_r(\Phi^{-1}(F_1(X_1)), \Phi^{-1}(F_2(X_2)))$ with Pearson product moment correlation r .	$\frac{\partial C}{\partial u_1}(u_1, u_2) = \Phi\left(\frac{\Phi^{-1}(u_2) - r\Phi^{-1}(u_1)}{\sqrt{1-r^2}}\right)$	$u_2 = \Phi(r\Phi^{-1}(u_1) + \sqrt{1-r^2}\Phi^{-1}(u_2))$ (2.4)
t-copula	$C_T(F_1(X_1), F_2(X_2)) = t_{r,v}(t_v^{-1}(F_1(X_1)), t_v^{-1}(F_2(X_2)))$ where t_v is the univariate student's t distribution function, with v degrees of freedom, and $t_{r,v}$ the bivariate distribution corresponding to t_v with Pearson product moment correlations r .	$\frac{\partial C}{\partial u_1}(u_1, u_2)$ $= t_{v+1}\left(\frac{\sqrt{v+1}}{\sqrt{v + [t_v^{-1}(u_1)]^2}} \frac{t_v^{-1}(u_2) - r t_v^{-1}(u_1)}{\sqrt{1-r^2}}\right)$	$u_2 = t_v(r t_v^{-1}(u_1) + \sqrt{1-r^2} \sqrt{\frac{v + [t_v^{-1}(u_1)]^2}{v+1}} t_{v+1}^{-1}(u_2))$ (2.5)

The normal copula is the copula that underlies the multivariate normal distribution. It shares the same dependency structure with the multivariate normal distribution; it uses pair-wise Pearson product moment correlations to measure dependency and allows arbitrary marginal distributions for the uncertainties. The normal copula has upper and lower tail dependencies equal to zero. The t-copula presents symmetric and positive upper and lower tail dependence which indicates a tendency for the t-copula to generate joint extreme events. When the number of degrees of freedom increases, the t-copula converges to the normal copula. For a limited number of degrees of freedom, however, the behaviors of the two copulas are quite different.

Archimedean copulas model upper tail dependency, lower tail dependency or both, so that they provide additional flexibility to describe the behavior of tail dependency in realistic situations. An Archimedean copula can be defined as follows:

$$C_A(u_1, u_2) = \varphi^{[-1]}(\varphi(u_1) + \varphi(u_2))$$

$C_A(u_1, u_2)$ is the Archimedean copula function with u_1 and u_2 as uniform variables, φ is the generator and $\varphi^{[-1]}$ is the inverse generator. The choice of the generator determines the Archimedean copula family. Frank, Clayton and Gumbel copulas are the most popular Archimedean copulas. The key formulae for these Archimedean copulas are summarized in Table 2 (Cherubini et al., 2004).

Table 2 Key Formulae for Archimedean Copulas

Family	Copula Description	Partial derivative of the copula	u_2
Clayton copula	$C_{Clayton}(u_1, u_2)$ $= (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}$	$\frac{\partial C}{\partial u_1}(u_1, u_2)$ $= u_1^{-\theta-1}(u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}-1}$	$u_2 = ((\alpha_2^{-\frac{\theta}{1+\theta}} - 1)\alpha_1^{-\theta} + 1)^{-\frac{1}{\theta}}$ (2.6)
Gumbel copula	$C_{Gumbel}(u_1, u_2)$ $= \exp\{-[(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{\frac{1}{\theta}}\}$	$\frac{\partial C}{\partial u_1}(u_1, u_2) = \frac{\varphi^{-1(1)}(c_2)}{\varphi^{-1(1)}(c_1)}$ with $c_1 = \varphi(u_1) = (-\ln(u_1))^\theta$ $c_2 = \varphi(u_1) + \varphi(u_2)$ $= (-\ln(u_1))^\theta + (-\ln(u_2))^\theta$ and $\varphi^{-1(1)}(t) = -\frac{1}{\theta} e^{-t^{1/\theta}} (t^{1/\theta})^{1-\theta}$	This equation has to be numerically solved with respect to u_2 . (2.7)
Frank copula	$C_{Frank}(u_1, u_2)$ $= -\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right)$	$\frac{\partial C}{\partial u_1}(u_1, u_2)$ $= \frac{(e^{-\theta u_2} - 1)e^{-\theta u_1}}{(e^{-\theta} - 1) + (e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}$	$u_2 = -\frac{1}{\theta} \ln\left(\frac{(e^{-\theta} - 1)\alpha_2}{e^{-\theta\alpha_1} - (e^{-\theta\alpha_1} - 1)\alpha_2} + 1\right)$ (2.8)

The Clayton copula (Clayton 1978) exhibits asymmetric lower tail dependence and is best suited for applications in which two outcomes are likely to experience low values together. For instance, Accioly and Chiyoshi (2004) explored the relationship between the drilling duration and the measured depth of the oil reservoir development, which shows clear lower tail dependence. They suggested the use of the Clayton copula to model the dependency between these two uncertainties. The Gumbel copula (Gumbel 1960) exhibits asymmetric upper tail dependence and is best suited for applications in which two outcomes are likely to experience high values together such as the performance of stock returns during a market jump. The Frank copula (Frank 1979) exhibits symmetric dependence in both tails. However, compared to the normal copula, the dependence in the Frank copula is weaker in both tails, and stronger in the center of the distribution. This suggests that the Frank copula is best suited for applications in which tail dependence is relatively weak. Yi and Bier (1998) discussed an example of the use of the Frank copula for accident frequencies and accident precursor in precursor analysis. Other examples include the use of Archimedean copulas models for aggregating expert opinions (Jouini and Clemen, 1996). A good introduction and summary of copulas can be found in Nelsen (1999).

3. The Dependent Decision Tree Approach

We now describe a general procedure for creating a dependent decision tree using any of the copula families listed above. There are four steps in the dependent decision tree approach.

Step 1. Assessment of marginals, dependence and copula. We start with the input information assessments: marginal distributions, measures of dependency, and the choice of a copula. Different techniques have been proposed in the literature for choosing the appropriate copula based on an analysis of historical data, as discussed by Yi and Bier (1998), Nelson (1999), Cherubini et al. (2004), Accioly and Chiyoshi (2004), Kotz and van Dorp (2010) and others. When these approaches are used in the context of decision analysis, the choice of the copula may also reflect subjective judgments regarding the existence of tail dependence between the uncertainties, and these judgments may guide the selection of an appropriate Archimedean copula as discussed by Armstrong, et al. (2004) and by Jouini and Clemen (1996).

If an elliptical copula is used, we can estimate the Spearman or Kendall's rank order correlations or the Pearson product moment correlations as the input correlation structure. These correlations may be determined from paired data sets using statistical approaches. However, in traditional decision analysis problems, these dependence measures may have to be assessed subjectively. C&R (1999) and Clemen et al. (2000) discuss subjective correlation assessment methods including estimating the probability of concordance and conditional fractile estimates. Based on our own experiences, we have found that subjects may be able to provide more accurate estimates of correlations than of conditional probabilities, where subjects making the latter assessments may confuse the notions of causation and correlation (Wright, et al., 1994). We refer the interested readers to these references for a detailed discussion of practical correlation assessment methods.

Step 2. Specification of parameters for the underlying copula. Since we are modeling the dependency structure through the use of a copula, we need to estimate the parameters for the specified copula. For the elliptical copulas, this is based on an estimate of the correlation between the original uncertainties.

The normal copula is parameterized in terms of the product moment correlation r^* . We distinguish the correlation for the copula with the superscript *, and write the correlation determined in Step 1 for the original random variables with no superscript. Intuitively, because the copula uniquely determines the dependency structure, there is a one to one relationship

between the product moment correlation r^* for the normal copula function and the correlation for the original variables.

The relationship between the product moment correlation r^* for the normal copula and the Spearman's rank order correlation ρ or Kendall's rank order correlation τ for the original uncertainties may be determined from a simple formula. For a specified Spearman's rank order correlation ρ , the formula is

$$r^* = 2\sin\left(\frac{\pi\rho}{6}\right) \quad (3.1)$$

for a specified Kendall's rank order correlation τ , the formula is

$$r^* = \sin\left(\frac{\pi\tau}{2}\right) \quad (3.2) \quad (\text{Kruskal 1958}).$$

The Pearson product moment correlation r between the original uncertainties may also be used to determine the normal copula correlation through a correlation matching technique developed by Cario and Nelson (1997), and discussed in the Appendix 1. The analytical relation (3.2) between the Kendall's rank order correlation τ and the product moment correlation r^* holds more generally for the t-copulas (Hult and Lindskog 2002).

For Archimedean copulas, the correlations of the copulas are determined by the parameter θ . Nelson (1999) documented the relationships between the types of Archimedean copulas and Kendall's tau, and he showed that the value of θ can be easily determined through these relationships. These relationships are summarized in the Table 3 (Cherubini et al., 2004). For example, for the Clayton family of copulas, we would have the expression $\tau = \frac{\theta}{\theta+2}$ which is easily solved for the value of θ given the value of Kendall's τ .

Table 3 Archimedean copulas parameter calculation

Family	Range of θ	Kendall's τ
Clayton copula	$(0, \infty)$	$\frac{\theta}{\theta+2}$
Gumbel copula	$(1, \infty)$	$1 - \frac{1}{\theta}$
Frank copula	$(-\infty, \infty)$	$1 + \frac{4}{\theta}(D_1(\theta) - 1)$ with $D_1 = \frac{1}{\theta} \int_0^\theta \frac{t}{e^t - 1} dt$

Step 3. Construction of the transient tree structure for the underlying copula. With the specified parameters, the underlying copula is modeled as a probability tree with discrete approximations to the conditional probability distributions at the endpoints of each branch of the predecessor chance node. C&R (1999) used the extended Pearson Tukey method (Keefer and

Bodily, 1983) for the discrete approximation to the probability distributions, and we will do the same. The extended Pearson Tukey (EP-T) Method approximates a continuous distribution by a three-point discrete distribution with probabilities 0.185, 0.630, and 0.185 assigned to the 5th, 50th, and 95th percentiles of the continuous distribution. The robust performance of the EP-T method is discussed by Keefer and Bodily (1983) and Keefer (1994). However, it is important to note that other methods could be used to develop these discrete approximations in a straightforward manner.

We can generate the transient tree structure for the underlying copula in two steps. First, we build the discrete approximation for X_1 with the 5th, 50th and 95th percentiles according to the EP-T method, which corresponds to the realizations of 0.05, 0.5, and 0.95 for u_1 , the unconditional percentile of X_1 . Second, we discretize the conditional X_2 given the outcome of X_1 with the 5th, 50th and 95th percentiles for the conditional distribution according to the EP-T method, which provides the choices of the conditional percentiles α_2 . These conditional percentiles α_2 and the realizations of the unconditional percentiles u_1 are then used to compute the dependent uniform variable u_2 through the formulas listed in Tables 1 and 2. Notice that the marginal distributions of X_1 and X_2 are not used; X_1 and X_2 are purely auxiliary in constructing the transient underlying copula tree.

Figure 1 Normal copula tree

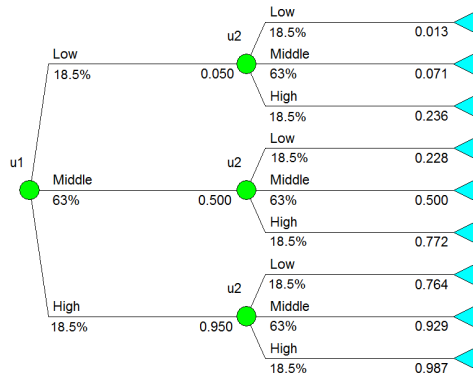


Figure 1 illustrates the constructed transient normal copula tree for the bivariate uniform variables u_1 and u_2 with Kendall's rank order correlation 0.7. This tree was constructed as follows: First, u_1 was approximated with three discrete points 0.05, 0.5 and 0.95 according to the EP-T method. Second, the dependent uniform variable u_2 given each realization of u_1 and the conditional percentiles were calculated using (2.4) from Table 1. For example, if u_1 is 0.05, then the conditional chance node for u_2 is calculated as 0.013, 0.071, and 0.236 for the conditional 5th, 50th, and 95th percentiles, respectively. The conditional relationship between the two uniform

uncertainties u_1 and u_2 is easy to see in the tree. For example, if the value of u_1 is high, then the values of the corresponding distribution of u_2 tend to be high, reflecting the positive correlation between these two variables.

Similarly, we can construct the corresponding transient trees for bivariate t, Frank, Clayton and Gumbel copulas using formulas (2.5)-(2.8). The same Kendall's rank order correlation of 0.7 was used to determine the values of the parameters for these copulas. Figures 2-5 show the constructed trees for these underlying copulas.

Figure 2 t-copula tree (d.f.=5)

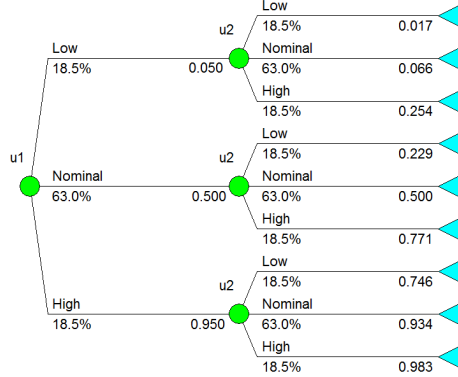


Figure 3 Clayton copula tree

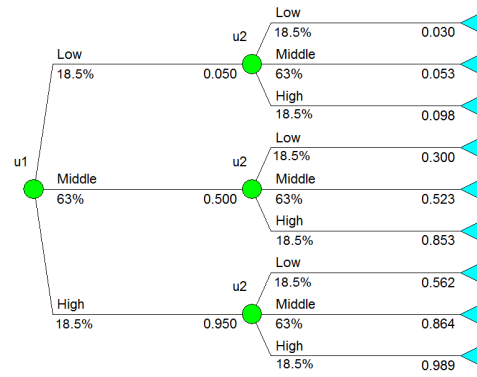


Figure 4 Frank copula tree

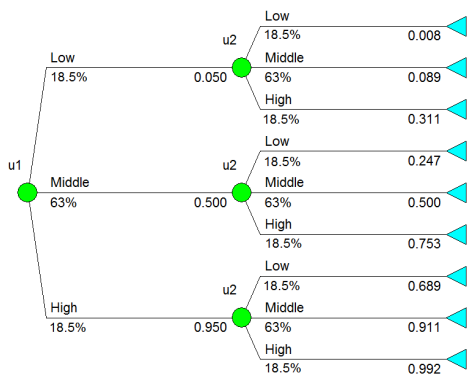
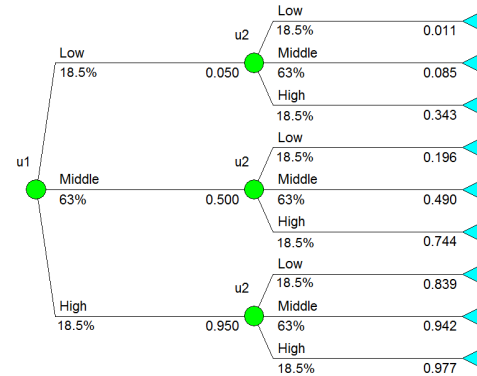


Figure 5 Gumbel copula tree



In contrast to the normal copula tree, the t-copula tree has a fatter tail dependence on both sides; the Frank copula tree has a thinner tail dependence on both sides; the Clayton copula tree has a fatter left tail dependence and thinner right tail dependence; the Gumbel copula tree has a fatter right tail dependence and thinner left tail dependence. These observations are consistent with the theory and provide different tail dependence structures between the uncertainties.

Step 4. Point-to-point inverse marginal transformation. After the calculation of u_1 and u_2 , we are ready to combine the marginal information and transform them to obtain the discrete approximations to the original uncertainties. The discrete approximations of X_1 and X_2 are obtained by applying the inverse of the target marginal distribution function for each realization of u_1 and u_2 : $x_1 = F_1^{-1}(u_1)$ and $x_2 =$

$F_2^{-1}(u_2)$. In the tree structure, this is a point-to-point transformation mapping the transient probability tree for the underlying copula into the tree structure for the original uncertainties. This component-wise inverse marginal transformation ensures that the resulting probability tree structure is for the target marginal uncertainties that were identified in Step 1.

We will use a bivariate case to illustrate this inverse transformation step. Let X_1 and X_2 be bivariate continuous variables with marginals $X_1 \sim N(8,1)$, $X_2 \sim \text{Gamma}(6,2)$, respectively, and with Kendall's rank order correlation 0.7. Based on the constructed underlying copula trees from Step 3, X_1 and X_2 are obtained by taking the inverse distribution for each realization of u_1 and u_2 in the tree structures. For example, X_1 is approximated with three discrete points 6.355, 8, and 9.645, the 5th, 50th, and 95th percentiles of the normal distribution with mean 8 and standard deviation 1; then, we implement the inverse gamma transformation for each realization of u_2 with the specified copulas calculated in Step 3. Consequently, for each specific copula, we model the variables X_1 and the conditional $X_2|X_1 = x_1$ with separate chance nodes in a probability tree.

Figure 6 Normal copula-based tree

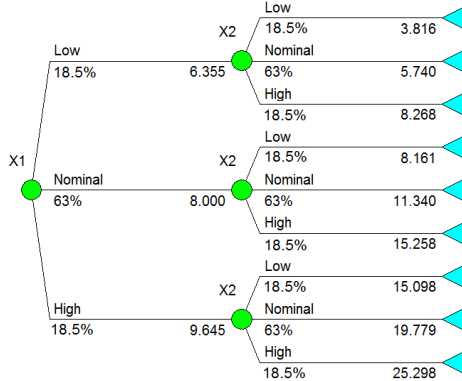


Figure 7 t-copula-based tree (d.f.=5)

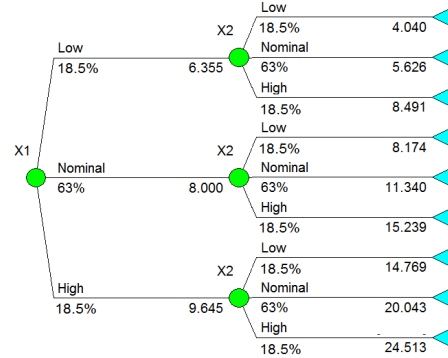


Figure 8 Frank copula-based tree

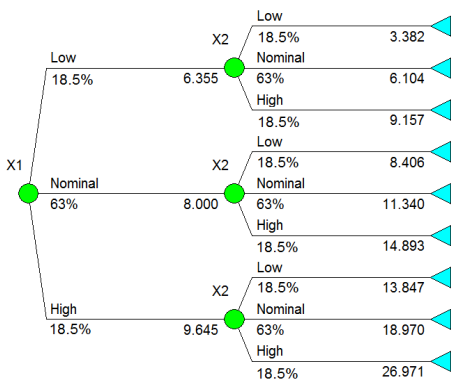


Figure 9 Clayton copula-based tree

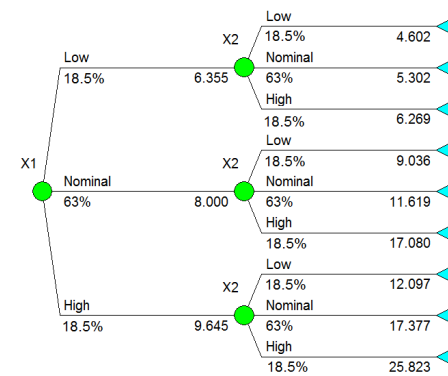
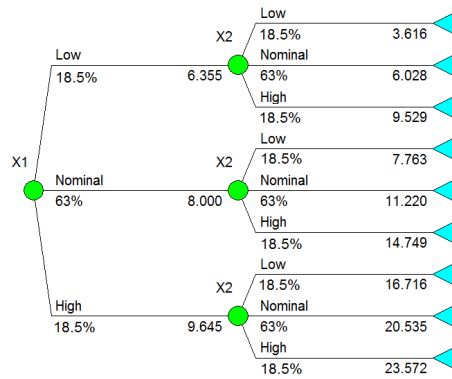


Figure 10 Gumbel copula-based tree



Figures 6-10 show the constructed copulas-based tree for each of the copulas presented in Step 3. In each of the trees, the first uncertainty is a normal distribution and the second uncertainty is a gamma distribution. The uncertainties evolve in the tree structure as a sequence of dependent uncertainties. The conditional relationship between the two uncertainties X_1 and X_2 is easy to see in the trees. For example, if the outcome for X_1 is high, then the conditional gamma probability distribution for X_2 tends to have higher values, reflecting the positive correlation between these two variables. Since an analytical expression for the conditional density may not be available, the dependent decision tree approach provides a discrete approximation for the conditional distributions which may be very useful for decision and risk analysis.

This method of constructing the bivariate copulas-based decision tree can be naturally extended to the multivariate copulas-based case. The formulae for the multivariate case corresponding to those in Tables 1 and 2 are shown in Table A1 in Appendix 2. In general, to build the multivariate copula-based decision tree for (X_1, \dots, X_n) , we first construct a discrete approximation for the unconditional uniform variable u_1 , then recursively compute the dependent uniform variables u_k ($k = 2, \dots, n$), conditioning on each of the point realizations of the previous discrete approximations for (u_1, \dots, u_{k-1}) . Taking the point-to-point inverse marginal transformation, each realization of (X_1, \dots, X_n) is discretely approximated using the extended Pearson Tukey method. Throughout the tree, the point realizations of these approximations vary with the conditioning scenarios; the probabilities assigned for the three branches of a chance node do not vary with the conditioning scenarios using the logic of the EP-T approximations.

The dependent probability trees constructed using this approach could be used for financial or risk analysis. Alternatively, they could be linked to other probability nodes representing independent uncertainties and to decision nodes to complete a dependent decision tree appropriate for the analysis of a decision problem.

4. Discussion of the C&R approach

C&R (1999) used the multivariate normal copula to capture the dependence structure among multiple random variables. The most attractive features of the multivariate normal copula are its flexibility and analytical tractability. Many multivariate copula families are constrained regarding how much dependency they can capture (Conway 1979). The class of multivariate Archimedean copulas, for example, is limited to intra-class correlation matrices (that is, each pair of uncertainties shares the same correlation) and can only model positive dependency for higher dimensions (Jouini and Clemen, 1996). In contrast, the multivariate normal copula allows the full range $[-1, 1]$ in pair wise correlations and is therefore a general and robust copula for most applications (Song 2000). Combining the multivariate normal copula with marginal distributions, a large variety of multivariate distributions can be produced in a unified fashion (C&R 1999; Avramidis et al. 2009).

Therefore, we will focus on applications of the multivariate normal copula in the next two sections, and show how the dependent decision tree approach that we have described provides computational advantages over the implementation used by C&R. In addition, we will introduce a novel application of the dependent decision tree using the multivariate normal copula to the evaluation of correlated cash flows in a manner that offers promise for the valuation of real options.

The dependent decision tree provides a computationally efficient alternative to the C&R approach. The major differences between the C&R approach and the proposed dependent decision tree approach are in Steps 3 and 4. Table 4 summarizes the two approaches.

Table 4. Comparison of the C&R approach and the Dependent Decision Tree approach

	The C&R approach	The Dependent decision tree approach
Step 1	Assessment of marginals, dependence and copula	The same
Step 2	Specification of parameters for the underlying copula	The same
Step 3	Construct the conditional density functions for the original uncertainties	Construct the transient tree structure for the underlying copula
Step 4	The continuous conditional density functions are discretized to construct the decision tree	Point-to-point inverse marginal transformation

Discussion of Step 3. The third step of the C&R approach is to construct the conditional density functions. With the normal copulas-based joint distribution $F(X_1, \dots, X_n)$, C&R (1999) show the corresponding joint density of these random variables can be obtained from

$$f(X_1, \dots, X_n) = f_1(X_1) \times \dots \times f_n(X_n) \times \exp\{-(\Phi^{-1}(F_1(X_1)), \dots, \Phi^{-1}(F_n(X_n))) \\ \times \frac{(\Sigma_Z^{-1} - I)}{2} \times (\Phi^{-1}(F_1(X_1)), \dots, \Phi^{-1}(F_n(X_n)))^T\} / |\Sigma_Z|^{1/2}$$

Furthermore, the conditional density function is calculated as

$$f(X_i|X_1, \dots, X_{i-1}) = \frac{f(X_1, \dots, X_i)}{f(X_1, \dots, X_{i-1})}, \quad i=2, \dots, n.$$

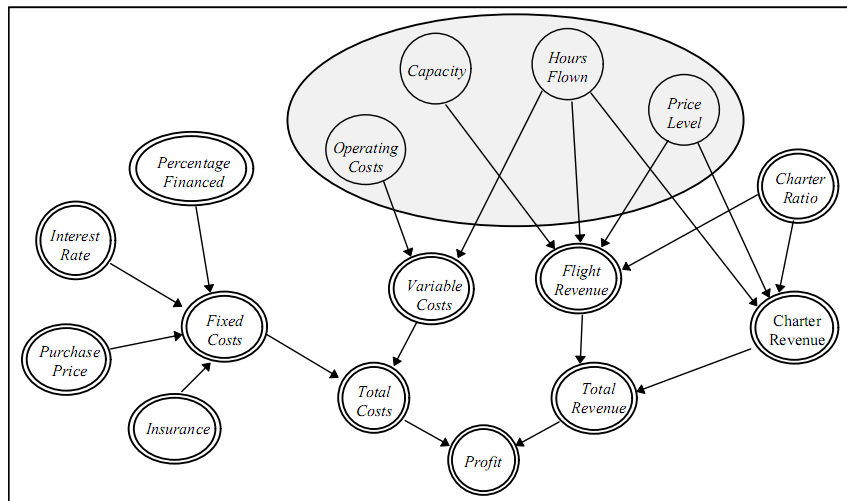
Instead of calculating the joint and conditional density functions of the original random variables, we construct only the discrete approximations of the underlying uniform distributions in the form of discrete probability distributions, using simple formulae.

Discussion of Step 4. The C&R approach derives the conditional density functions explicitly. The difficulty in obtaining the conditional density functions increases with the complexity and dimension of the model. Additionally, because the calculated conditional density functions usually do not belong to familiar distribution families, the C&R approach requires the additional computational cost of numerical integration or Monte Carlo simulation to calculate the percentiles used by the extended Pearson Tukey method for the discrete approximations of the chance nodes in the decision trees.

The C&R approach is computationally inefficient. As a discrete approximation of the probability model, we only need certain percentiles of the conditional density function to construct the decision trees. Therefore, the calculation of the complete conditional density function is unnecessary. The dependent decision tree approach is much more efficient because it calculates only the necessary percentiles of the conditional density function and established a point-to-point transformation to determine the target variables. This point-to-point transformation saves a significant computational cost.

Case Example: Eagle Airlines. Clemen and Reilly (1999, 2000) described an example of a decision model involving multiple correlated uncertainties. In this hypothetical decision analysis problem, the owner of Eagle Airlines is considering the best choice between purchasing a used aircraft outright, an option to purchase it within a year at a specified price, and a money market alternative investment. The influence diagram in Figure 11 portrays the initial model. The multiple critical variables are identified from a sensitivity analysis to be price, hours flown, capacity and operating cost, which are correlated. As the profit generated from purchasing the used aircraft is driven by these four critical variables, the decision model must reflect the marginal distribution of each variable, the correlations among the variables, and the flexibility of buying it later.

Figure 11 The Influence Diagram for the Eagle Airlines Problem



Source: Clemen and Reilly (1999)

C&R (1999) use this hypothetical decision faced by Eagle Airlines to illustrate the decision tree model constructed from the C&R approach. For comparison's sake, we will use the dependent decision tree approach to model the same example and discuss the computational advantage of the latter.

Step 1. Assessment of marginals, dependence and copula. We will denote the four critical variables, price, hours flown, capacity, and operating cost by P , H , C , and O , respectively, and use the same information as described in C&R (1999) for marginals, correlation and the use of the normal copula. The probability tree will be constructed based on sequencing the chance nodes in this order. The marginal densities for the four variables are indicated in Table 5. The dependencies among the four variables are measured by the Spearman's rank order correlations shown in Table 6.

Table 5 Marginal Distributions for the Eagle Airlines Probability Model

Variable	Distribution	Parameters		Range
Price Level (P)	Scaled Beta	$\alpha=9$	$\beta=15$	[\$81.94, \$133.96]
Hours Flown (H)	Scaled Beta	$\alpha=4$	$\beta=2$	[66.91,1135.26]
Capacity (C)	Beta	$\alpha=20$	$\beta=20$	[0,1]
Operating Cost (O)	Normal	$\mu=245$	$\sigma=11.72$	$(-\infty, +\infty)$

Table 6 Spearman correlation Σ_X

	P	H	C	O
P	1			
H	-0.5	1		
C	-0.25	0.5	1	
O	0	0	0.25	1

Step 2. Specification of parameters for the underlying copula. To construct the dependent decision

tree model for Eagle Airlines, we first need to specify the product moment correlation matrix Σ_Z for the underlying normal copula. Since all variables are continuous and the specified correlation structure is the Spearman's rank order correlation, we can apply the transforming formula (3.1) $r_{ij}^* = 2\sin(\frac{\pi\rho_{ij}}{6})$ for $i, j \in \{P, H, C, O\}$ to calculate the corresponding product moment correlation matrix Σ_Z shown in Table 7.

Table 7 Correlation Matrix Σ_Z

	<i>P</i>	<i>H</i>	<i>C</i>	<i>O</i>
<i>P</i>	1			
<i>H</i>	-0.5176	1		
<i>C</i>	-0.2611	0.5176	1	
<i>O</i>	0	0	0.2611	1

Step 3. Construction of the transient tree structure for the underlying copula. The third step is to generate a normal copula based transient probability tree for uniform variables (u_P, u_H, u_C, u_O) .

To create the probability tree for the standardized uniform variables, we first generate the extended Pearson Tukey discretization for u_P . It is a three point discrete approximation for the standard normal distribution with probabilities 0.185, 0.630, and 0.185 assigned to the 0.05, 0.5 and 0.95. The subsequent discrete chance nodes are contingent on the outcomes of the precedent nodes. We apply the Cholesky factorization to decompose Σ_Z into the lower triangular Cholesky matrix shown in Table 8 to assist the calculation of the dependent uniform variables. These calculations are adapted from the NORTA approach developed by Cario and Nelson (1997) to support high dimensional Monte Carlo simulation and offer computational advantages relative to the C&R approach, as we will discuss (see details in Appendix 1 and Table A1 in Appendix 2).

Table 8 Decomposed Lower Triangular Cholesky Matrix

Cholesky	<i>P</i>	<i>H</i>	<i>C</i>	<i>O</i>
<i>P</i>	1	0	0	0
<i>H</i>	-0.5176	0.8556	0	0
<i>C</i>	-0.2610	0.4470	0.8555	0
<i>O</i>	0	0	0.3051	0.9523

There are three dependent uniform variables to calculate:

(1) The dependent uniform u_H given the outcomes of u_P . Using formula (2.4) for the bivariate case, we can calculate u_H as follows:

$$u_H = \Phi(-0.5176\Phi^{-1}(\alpha_1) + 0.8556\Phi^{-1}(\alpha_2))$$

(2) The dependent uniform u_C given the outcomes of u_P and u_H . Using multivariate version of the formula given in Appendix 2, u_C is calculated as follows:

$$u_C = \Phi(-0.2610\Phi^{-1}(\alpha_1) + 0.4470\Phi^{-1}(\alpha_2) + 0.8555\Phi^{-1}(\alpha_3))$$

(3) Similarly, the dependent uniform u_O given the outcome of u_P , u_H , and u_C is calculated as:

$$u_O = \Phi(0.3051\Phi^{-1}(\alpha_3) + 0.9523\Phi^{-1}(\alpha_4))$$

For instance, if the outcome of u_P is 0.05, then the conditional chance node for $u_H|u_P = 0.05$ is calculated for the 5th, 50th, and 95th percentiles, and the three contingent outcomes are determined to be 0.289, 0.803, and 0.988 respectively. Similarly, we create the contingent tree for each successive node until we generate the complete multivariate standard decision tree for (u_P, u_H, u_C, u_O) .

Step 4. Point-to-point inverse marginal transformation. The last step is the use of the inverse marginal transformations to transform the discrete approximation for the multivariate standard uniform random vector (u_P, u_H, u_C, u_O) into the corresponding discrete approximation for the original random vector X .

$$X = (P, H, C, O) = (F_P^{-1}[u_P], F_H^{-1}[u_H], F_C^{-1}[u_C], F_O^{-1}[u_O])$$

where F_P, F_H, F_C, F_O are the marginal distribution functions of the variables P, H, C, O respectively. For example, the three outcomes of the chance node representing P are $F_P^{-1}[0.05] = 93.469$, $F_P^{-1}[0.5] = 101.264$, and $F_P^{-1}[0.95] = 110.055$. Given the three calculated conditional outcomes of $u_H|u_P = 0.05$, 0.289, 0.803, and 0.988 in the transient normal copula tree, the corresponding conditional outcomes of H given $P = 93.469$ are $F_H^{-1}[0.289] = 677.220$, $F_H^{-1}[0.803] = 956.638$, and $F_H^{-1}[0.988] = 1096.970$. These results are identical to those obtained for the Eagle airlines problem by using the C&R approach. Since the inverse functions of beta and normal distributions are available in Excel, the multivariate probability tree can be easily constructed in the Excel spreadsheet. The constructed dependent probability tree can be augmented with appropriate decision nodes to create a dependent decision tree to be used for the analysis of the decision problem.

The construction of the dependent decision tree adapted from the NORTA approach for the normal copula requires only simple calculations that can be implemented in Excel with spreadsheet formulas. As a comparison, the C&R approach requires the calculation of the conditional density functions and their percentiles for discretization. Let $f_H(h)$, $f_C(c)$, $f_O(o)$ denote the marginal density function of the variables H, C, O respectively. The C&R approach calculates the conditional density functions for H, C , and O as follows (C&R, 1999):

$$f(h|p) = f_H(h) \exp\{-0.5[\Phi^{-1}[F_H(h)] + 0.518\Phi^{-1}[F_P(p)]^2/0.732 - (\Phi^{-1}[F_H(h)]^2)/0.732^{1/2}$$

$$f(c|p, h) = f_C(c) \exp\{-0.5[\Phi^{-1}[F_C(c)] - 0.009\Phi^{-1}[F_P(p)] - 0.523\Phi^{-1}[F_H(h)]^2/0.732$$

$$- (\Phi^{-1}[F_C(c)]^2)/0.732^{1/2}$$

$$f(o|p, h, c) = f_o(o) \exp\{-0.5[\Phi^{-1}[F_o(o)] + 0.003\Phi^{-1}[F_p(p)] + 0.186\Phi^{-1}[F_h(h)] \\ - 0.357(\Phi^{-1}[F_c(c)]^2)/0.907 - (\Phi^{-1}[F_o(o)]^2)/0.907^{1/2}\}$$

The computational complexity required to obtain these conditional density functions is significant and the procedure is difficult to standardize. Additionally, because none of these conditional density functions are familiar distributions, assessment methods such as Monte Carlo simulations have to be used to calculate the 5th, 50th, and 95th percentiles for each conditional distribution required by the extended Pearson Tukey method for the branch level of each factor.

5. Extension: The Consolidated Dependent Decision Tree

In the previous example, the dependent decision tree included multiple uncertainties from the same time period. However, this dependent decision tree framework can be naturally extended in a multivariate time series context to construct a probability tree for sequential, correlated uncertainties. A sequential probability tree is appropriate with information evolves dynamically and when decisions are made sequentially.

Suppose we have a problem with n correlated random variables in each time period $1, \dots, T$ that determine a payoff or cash flow in each of these periods. Such a problem can be modeled with nT correlated uncertainties in a decision tree. While there could be situations in which this is advantageous (for instance, multi-asset path-dependent options), it will increase the dimension of the problem and require more computational cost due to the corresponding tree size. When the payoff is a function of the realization of multiple uncertainties in each period, we can use Monte Carlo simulation to consolidate the multiple uncertainties into one single uncertainty in each period in order to reduce the size of the constructed decision tree.

In this section, we will discuss the integration of the dependent decision tree approach and Monte Carlo simulation to construct a consolidated dependent decision tree for the multivariate time series. The consolidated dependent decision tree handles multiple factors of uncertainty simultaneously in the simulation of the project cash flow to estimate the marginal and the autocorrelation structure for the project value, and then applies a dependent decision tree approximation to the consolidated project cash flow. This allows the decision nodes introduced into the dependent decision tree to conveniently model sequential decisions that represent managerial flexibility and that may impact this cash flow, such as a decision to abandon the project before the last time period.

We now address the implementation of this extension.

Step 0. Consolidation of uncertainties. As a prerequisite step, the consolidated extension uses Monte Carlo simulation to aggregate a multivariate time series process into a consolidated single-factor auto-correlated time series process based on the project cash flow. This simulation may be implemented in a spreadsheet with a standard add-in such as @Risk or Crystal Ball.

Step 1. Assessment of marginals, dependencies and copula. With the simulated data for the consolidated project cash flow, we can assess the auto-correlation structure, the marginal distributions and appropriate copula for the project cash flow in each period. We can either fit the simulated project cash flow with some standard parametric family of distributions, or with a more flexible alternative, the Johnson translation system that matches the first four moments of the simulated data (Swain et al. 1988). When there is no familiar parametric distribution associated with the simulated project cash flow, the use of the Johnson translation system to characterize the marginal distributions of the simulated data provides standardization and more flexibility in modeling uncertainties. Moreover, it offers additional computational convenience for generating dependent decision trees, as we discuss in Step 4.

The Johnson translation system for a random variable X is defined by a distribution function of the form

$$F_X(x) = \Phi \left\{ \gamma + \delta f \left[\frac{x - \xi}{\lambda} \right] \right\},$$

where γ and δ are shape parameters, ξ is a location parameter, λ is a scale parameter, and $f(\cdot)$ is one of the following transformations:

$$f(y) = \begin{cases} \log(y), & \text{for the } S_L(\text{lognormal}) \text{ family} \\ \log(y + \sqrt{y^2 + 1}), & \text{for the } S_U(\text{unbounded}) \text{ family} \\ \log\left(\frac{y}{1 - y}\right), & \text{for the } S_B(\text{bounded}) \text{ family} \\ y, & \text{for the } S_N(\text{normal}) \text{ family} \end{cases}$$

There is a unique family (choice of f) for each feasible combination of the skewness and the kurtosis that determine the parameters γ and δ . Any mean and variance can be attained by any one of the families by the manipulation of the parameters λ and ξ . Within each family, a distribution is completely specified by the values of the parameters $[\gamma, \delta, \lambda, \xi]$ and the range of X depends on the family of interest. A robust and computationally efficient procedure for identifying the type of transformation and estimating Johnson-type marginals is suggested by Swain et al. (1988) and implemented in software called FITTR1. A detailed illustration for the Johnson-type probability density function can be found in Biller and Nelson (2005).

Steps 2 and 3 are identical to the corresponding steps in the dependent decision tree approach.

Step 4. Point-to-point inverse transformation. The Johnson system may be used to give the approximated continuous marginal density for the inverse marginal transformation.

With the Johnson system, the evaluation of the composite function $F_X^{-1}(u)$ is simplified significantly because $F_X^{-1}(u) = \xi + \lambda f^{-1}\left[\frac{\Phi^{-1}(u) - \gamma}{\delta}\right]$, where

$$f^{-1}[a] = \begin{cases} \exp(a), & \text{for the } S_L(\text{lognormal}) \text{ family} \\ \frac{\exp(a) - \exp(-a)}{2}, & \text{for the } S_U(\text{unbounded}) \text{ family} \\ \frac{1}{1 + \exp(-a)}, & \text{for the } S_B(\text{bounded}) \text{ family} \\ a, & \text{for the } S_N(\text{normal}) \text{ family} \end{cases}$$

Case Example: Oil Production Example. We use the hypothetical oil production project discussed by Brandão et al. (2005a, 2005b) and Smith (2005) to illustrate the consolidated dependent decision tree extension. In each year t ($t=1, \dots, 10$), the two relevant sources of uncertainty are the oil price, and the variable operation cost. The oil price process follows a Geometric Brownian Motion process with mean drift (rate of change) of 0% and volatility of 15% and the variable cost process follows a Geometric Brownian Motion process with mean drift of 2% and volatility of 10% under the fully risk neutral approach (Smith, 2005). The initial variable operating cost is \$10 per barrel and the initial oil price is \$25 per barrel. There is also a \$5 million per year fixed cost. The risk free interest rate is 5%. Table 9 presents the expected values of the future cash flow. All values are in millions of dollars. The present value of the expected cash flow is \$392 million.

Table 9 Base Case Expected Cash Flow for the Project

Year	1	2	3	4	5	6	7	8	9	10
Remaining Reserves	90.0	81.0	73.4	66.8	61.3	56.6	52.6	49.2	46.3	43.9
Production Level	9.0	7.7	6.5	5.5	4.7	4.0	3.4	2.9	2.5	2.1
Variable Operating Cost Rate	10.2	10.4	10.6	10.8	11.0	11.3	11.5	11.7	12.0	12.2
Oil Price	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0
Revenues	225.0	191.3	162.6	138.2	117.5	99.8	84.9	72.1	61.3	52.1
Production Cost	(96.8)	(84.6)	(74.0)	(64.8)	(56.9)	(50.0)	(44.0)	(38.8)	(34.3)	(30.4)
Cash Flow	128.2	106.7	88.6	73.4	60.6	49.9	40.9	33.3	27.0	21.7
Profit Sharing	(32.1)	(26.7)	(22.1)	(18.3)	(15.1)	(12.5)	(10.2)	(8.3)	(6.8)	(5.4)
Net Cash Flow	96.2	80.0	66.4	55.0	45.4	37.4	30.7	25.0	20.3	16.3

The net cash flow is a function of the two uncertainties: the oil price and the variable production cost. Instead of modeling each uncertainty as a separate chance node in each period in the decision tree, we

will model the net cash flow as a consolidated uncertainty $X_t, t = 1, \dots, 10$, in each period. As a result, we reduce the 20 uncertainties into 10 uncertainties in the constructed decision tree. The present value of the oil production project without options is the expectation of the total discounted net cash flow in the future. The net cash flow is a sequence of auto-correlated time series variables (X_1, \dots, X_{10}) . The net cash flow in a later time period is dependent on the outcomes of the previous net cash flows. Hence, we will model the project value without options as a 10-dimensional dependent decision tree. We use the multivariate normal copula to model the dependence structure between period to period cash flows.

First, we simulate the time series of net cash flow associated with both uncertainties. After running a large number (5,000) of iterations, the Monte Carlo simulation provides approximations to the distributions of the net cash flow for each period. Then, we use the Johnson translation system to characterize the marginal distributions for (X_1, \dots, X_{10}) and obtain the estimated distribution families and parameters $[\gamma, \delta, \lambda, \xi]$ for (X_1, \dots, X_{10}) shown in Table 10.

Table 10 Johnson Translation Parameters for Each Period										
Year	1	2	3	4	5	6	7	8	9	10
Distribution	S_B	S_B	S_U	S_U	S_U	S_U	S_U	S_U	S_U	S_U
γ	10.41	8.803	-3.383	-1.323	-1.569	-2.037	-3.519	-1.928	-1.395	-3.446
δ	5.425	4.029	3.284	2.053	2.086	2.187	2.484	2.073	1.811	2.332
λ	1264	1389	64.97	49.06	44.71	39.23	28.11	30.74	26.24	18.33
ξ	-67.53	-63.51	-16.82	16.97	4.141	-9.23	-28.5	-11.98	-5.747	-25.52

For example, X_1 , the cash flow of period 1, is categorized as the Johnson bounded S_B marginal distribution with parameters **[10.41, 5.425, 1264, -67.53]**.

$$F_{X_1}^{-1}(u_1) = -67.53 + 1264 / (1 + \exp((10.41 - \Phi^{-1}(u_1)) / 5.425))$$

And X_3 , the cash flow of period three, is categorized as the Johnson unbounded S_U marginal distribution with parameters **[-3.383, 3.284, 64.97, -16.82]**.

$$F_{X_3}^{-1}(u_3) = -16.82 + 64.97 \frac{\exp(\frac{3.383 + \Phi^{-1}(u_3)}{3.284}) - \exp(\frac{-3.383 - \Phi^{-1}(u_3)}{3.284})}{2}$$

Notice that the net cash flow from period 3 to period 10 are all categorized as the Johnson unbounded (S_U) marginal distribution, which has the capability of capturing the tail behavior of a wide variety of distributional shapes.

Next, we need to measure the dependencies among the variables. The decision maker can choose to match either the Pearson product moment correlation or the Spearman rank order correlation as the desired correlation structure. For simplicity, we use the Spearman rank order correlations in our illustration. The Spearman rank order correlations of (X_1, \dots, X_{10}) were calculated from the simulated

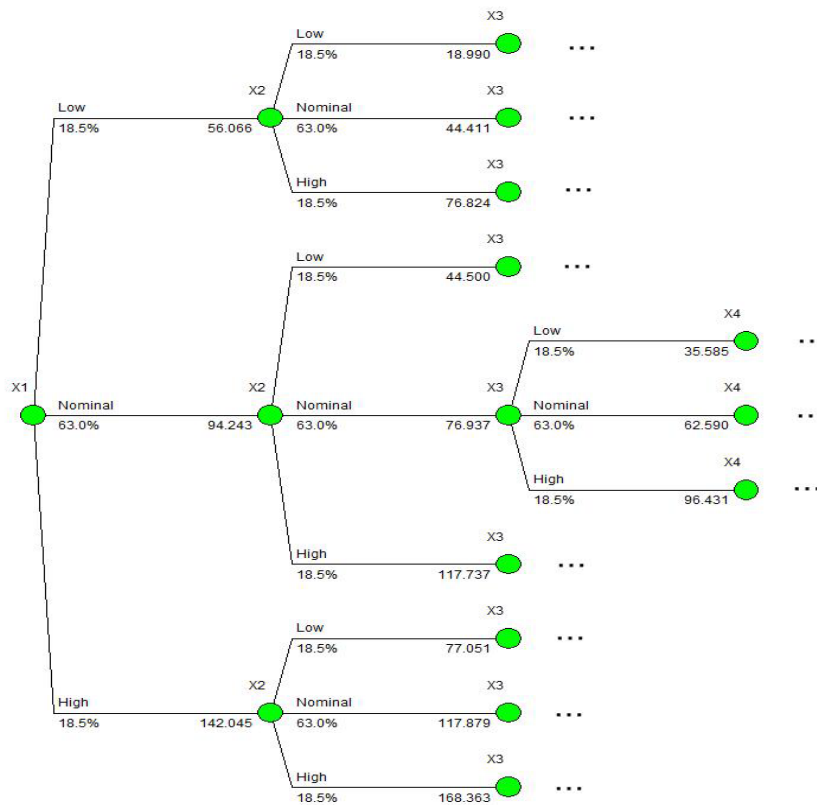
data, and the calculated corresponding product moment correlation matrix Σ_Z for the normal copula is reported in Table 11.

Table 11 The Transient Correlation Σ_Z

Year	1	2	3	4	5	6	7	8	9	10
1	1									
2	0.6913	1								
3	0.5712	0.7982	1							
4	0.4883	0.6828	0.8525	1						
5	0.4274	0.6008	0.7499	0.8805	1					
6	0.3781	0.5362	0.6751	0.7947	0.9022	1				
7	0.3427	0.4838	0.6114	0.7269	0.8230	0.9143	1			
8	0.3207	0.4578	0.5759	0.6831	0.7672	0.8513	0.9284	1		
9	0.2966	0.4245	0.5388	0.6375	0.7164	0.7925	0.8672	0.9350	1	
10	0.2784	0.3962	0.5050	0.5987	0.6767	0.7440	0.8154	0.8804	0.9423	1

Figure 12 shows a portion of the resulting consolidated dependent decision tree. For example, the net cash flow in year 1, X_1 , is modeled as a three branch discrete chance node with outcomes 56.066, 94.243, and 142.045, and with probabilities 18.5%, 63%, and 18.5% assigned to each corresponding outcome as before. The net cash flow in year 2, X_2 , is a conditional chance node. If the outcome of X_1 is 56.066, then the conditional net cash flow in year 2 is modeled as a three branch discrete chance node with outcomes 18.990, 44.411, and 76.824. The relationship between the variables reflects a positive auto-correlation. For example, if the outcome of X_1 is high the outcomes associated with X_2 tend to be high, and vice versa.

Figure 12 Consolidated Dependent Decision Tree for the Oil Production Example



The complete decision tree is available on request from the authors.

We compare the mean, median and standard deviation of the net cash flow in the consolidated dependent decision tree model with the statistics calculated from the simulated data in Table 12. They show that this constructed multivariate decision tree provides a close approximation to the target multivariate variables.

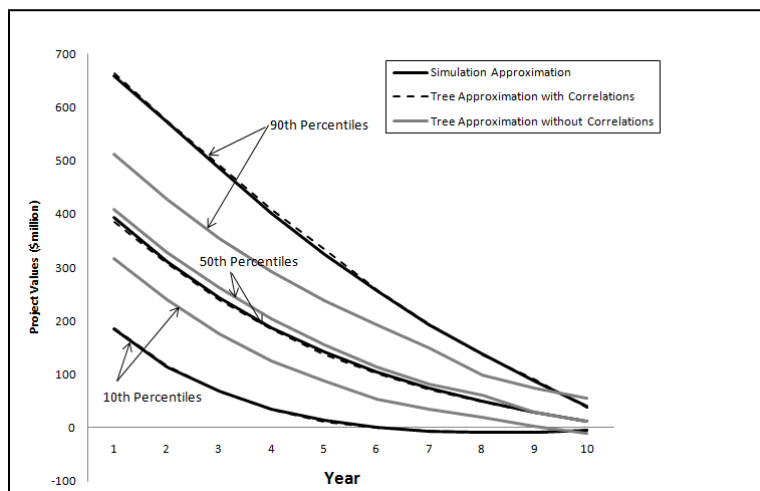
Table 12 Statistics of the Multivariate Decision Tree and of the Simulated Data

Cash Flow	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Decision Tree										
mean	96.02	80.04	66.36	55.08	45.52	37.46	30.65	24.97	20.19	16.23
median	94.24	76.94	62.59	50.82	41.03	32.44	26.00	20.90	16.65	12.68
standard deviation	26.25	32.04	33.14	33.33	31.93	30.04	27.65	25.21	22.95	20.65
Simulation										
mean	96.15	80.09	66.49	55.12	45.59	37.53	30.75	24.98	20.16	16.21
median	94.86	77.37	62.23	50.37	40.61	32.45	25.43	20.73	16.42	12.79
standard deviation	26.40	32.53	34.00	33.68	32.24	30.34	28.10	25.19	22.47	20.35

The evolution of the future project value reflects the auto-correlated cash flow and is meaningful for valuing managerial flexibilities associated with the underlying asset. Figure 13 shows the 10th, 50th, and 90th percentiles of the future project values using the Monte Carlo simulation model (in heavy lines), the

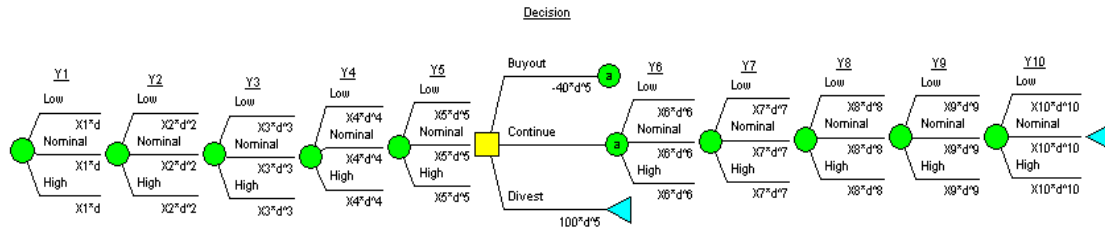
consolidated dependent decision tree model (in dashed lines) and the multivariate decision tree model if we ignore the auto-correlations (with light lines). We see that the consolidated dependent decision tree model provides a close approximation to the project values in each time period compared to the Monte Carlo simulation results in the oil production example. In contrast, ignoring the correlations will greatly underestimate the uncertainties in the future project values: the 90th percentiles are much too low and the 10th percentiles are much too high. Because uncertainty plays the key role in real options valuation, the underestimated uncertainties may lead to inaccurate real options value because of the poor approximation to the project values by ignoring the correlations between period to period cash flows.

Figure 13 Approximation of the Project Value in the Oil Production Example



At the end of year 5 in this example, there exist three alternatives for managerial flexibility: continuing the project, buying out the partner's 25% share for \$40 million, or selling the decision maker's share for \$100 million (divesting). We insert the decision nodes for the options in year 5 in the established decision tree as shown in Figure 14. The project value is estimated to be \$420.27 million for the dependent decision tree. Therefore, the estimated value of the options to buy out or divest is worth $\$420.27 - \$392 = \$28.27$ million which is within 2.5% of the \$29 million estimate obtained with the Longstaff and Schwartz Monte Carlo simulation approach (Smith, 2005). In contrast, the independent cash flow decision tree model gives an estimate of the project with these options of \$392 million, which implies no value at all for the buyout and divest options.

Figure 14 Decision Tree for Oil Production Example



Brandão et al. (2005a, 2005b) discussed the use of a binomial decision tree to solve this same problem. Following Copeland and Antikarov (2001), they estimated the parameters of the binomial decision tree by assuming that the change in the value of the oil production example over time could be closely approximated by a Geometric Brownian Motion process. While this assumption provided a reasonable approximation for this example, Smith (2005) noted that it might not be appropriate in general when used to model consolidated cash flow. Recently, Wang and Dyer (2010) suggested an implied binomial tree approach using risk-neutral probabilities inferred from the simulated cash flow information and solved this same example problem. While the implied binomial tree approach does provide more flexibility in modeling cash flow than the assumption of a Geometric Brownian Motion process, it also assumes that the underlying stochastic process is relatively “smooth” over time.

In contrast, the consolidated dependent decision tree begins with empirically determined approximations to the marginal distributions of the cash flow in each period, and to the correlations among the time periods. Therefore, it is a much more general approximation method that is not subject to limiting assumptions regarding the underlying stochastic process.

The Longstaff and Schwartz Monte Carlo simulation provides a close estimate of the option value for this case and was used as our benchmark; however, it may be difficult to implement for more complex real options problems. For a discussion of the relative benefits of using the simulation approach versus a decision tree approach for real options problems, see Brandão et al. (2005a, 2005b), Smith (2005) and Wang and Dyer (2010).

6. Discussion and Conclusion

In this paper, we proposed the dependent decision tree approach to construct decision trees for dependent multivariate random variables. In essence, the dependent decision tree approach constructs a probability tree for discrete approximations of the underlying copula with uniform variables and then inverse transforms it into the desired probability tree for the original multivariate random variables with the target marginal distributions and correlation structure. Decision nodes or other probability nodes

representing independent uncertainties may then be added to complete a decision tree for a specific application.

This approach is a general one, and we have illustrated how it can be applied with copulas from the elliptical and Archimedean families. The elliptical family includes the normal and the t-copulas, and the latter allows the construction of dependent decision trees consistent with joint distributions that have “fatter tails” than those associated with the normal copula. The Archimedean family includes several families of copulas that allow the construction of dependent decision trees that include tail dependency, and that may be of critical importance for some applications of discrete approximations in risk analysis and insurance, for example.

Because of its flexibility and analytical tractability, we have emphasized the use of the dependent decision tree with the normal copula, and contrasted this approach with the computational requirements of a similar method proposed by Clemen and Reilly (1999). The dependent decision tree approach and the C&R approach both specify separately the marginal distributions and dependencies among them, and take advantage of the convenience of the normal copula. The major advantages of the dependent decision tree approach are its simplicity and standardization. The computational cost of the C&R approach is mainly in the calculation of the conditional density functions and the percentiles of these conditional distributions. These calculations may be complex and cannot be standardized because of the need to deal with arbitrary marginal distributions.

The dependent decision tree approach is adapted from the NORTA approach used for generating dependent multivariate random variables for Monte Carlo simulation, and transfers this problem into calculating the conditional structure of the underlying copula. This greatly simplifies the procedure. The main computational cost of the dependent decision tree approach is the correlation matching problem that can be skipped for rank order correlations when all marginal distributions are continuous, and that can be carried out with efficient numerical methods for product moment correlations. As a result, the dependent decision tree approach is computationally more efficient than the C&R approach and intuitively simpler for practical applications. The computational benefit of the dependent decision tree approach is especially significant when the marginal distributions of the component random variables are neither identical nor from the same family of distributions. Additionally, the dependent decision tree approach provides more flexibility in the choice of copulas in modeling the various dependence structures of the multivariate uncertainties, and therefore extends the scope of decision tree models available in the decision analysis literature.

As with the C&R approach, the dependent decision tree approach constructs non-recombined decision trees, and therefore is subject to the "curse of dimensionality" problem. The number of nodes and

branches grows exponentially with the number of variables. For multivariate time series, we presented a way to reduce the dimension of the problem when the payoff is a function of multiple uncertainties in each period. While we used the extended Pearson Tukey method for discretization of the probability distributions in this study, the proposed method can be easily applied with other discretization methods to generate chance nodes with an arbitrary number of branches.

Uncertainties are often dependent. The use of the dependent decision tree approach provides a practical alternative procedure to model dependent uncertainties in decision trees.

Acknowledgements

The authors thank the anonymous referees and the associate editor for their helpful suggestions and insights. Their comments helped to improve the paper substantially. The authors acknowledge support from the Center for Petroleum Asset Risk Management at The University of Texas at Austin.

Appendix 1. When the dependency structure is modeled by the normal copulas, the dependent decision tree approach extends NORTA (NORmal To Anything) (Cario and Nelson 1997), a simulation algorithm for generating multivariate random variables, to modeling dependence in decision trees.

With the multivariate normal copula modeling the dependence among the random variables, we can transform a multivariate normal random vector Z into the desired random vector X . If we define $Z = (Z_1, \dots, Z_n) \sim N(0, \Sigma_Z)$, we have the representation

$$X = (X_1, \dots, X_n) = (F_1^{-1}(\Phi(Z_1)), \dots, F_n^{-1}(\Phi(Z_n)))$$

Note that $U_i = \Phi(Z_i)$ is the uniform variable for the normal copula function. Hence the NORTA approach is theoretically equivalent to the dependent decision tree approach we discussed when the normal copula is used for modeling dependence.

The NORTA approach provides efficient algorithms to determine the normal copula correlations $\Sigma_Z = (r_{ij}^*)_{i,j=1}^n$ through a correlation matching technique when we use the Pearson product moment correlations $\Sigma_X = (r_{ij})_{i,j=1}^n$ as the input correlation structure. Cario and Nelson (1997) show each pair of (r_{ij}^*, r_{ij}) satisfies the following equations:

$$\begin{aligned} r_{ij} &= \text{corr}(X_i, X_j) = \text{corr}\left(F_i^{-1}(\Phi(Z_i)), F_j^{-1}(\Phi(Z_j))\right) \\ &= \iint \left(\frac{F_i^{-1}(\Phi(Z_i)) - \mu_i}{\sigma_i}\right) \left(\frac{F_j^{-1}(\Phi(Z_j)) - \mu_j}{\sigma_j}\right) \phi(z_i, z_j; r_{ij}^*) dz_i dz_j \end{aligned}$$

where μ_i and σ_i^2 are the mean and variance of the marginal distribution F_i respectively, and ϕ is the bivariate standard normal density function with correlation r_{ij}^* . Because of the symmetry of the correlation matrix and the perfect correlation of any variable with itself, there are $n(n-1)/2$ independent correlation matching equations for r_{ij}^* , $1 \leq i, j \leq n$.

Although analytic solutions are only available for specific cases (Li and Hammond, 1975), efficient numerical procedures have been developed to solve these equations. For example, Cario and Nelson (1997) use numerical root-finding methods combined with two dimensional numerical integration methods to evaluate the double integral function values. Chen (2001) proposes a Monte Carlo simulation-based retrospective approximation, using a stochastic root-finding algorithm to estimate the r_{ij}^* . Avramidis et al. (2009) provide a discussion of the case of discrete marginals.

Appendix 2. The formulas for the multivariate copulas corresponding to Tables 2 and 3 are shown below:

Table A1. Key Formulae for Multivariate Copulas

Family	Copula Description	
normal copula	$C_N(u_1, \dots, u_n) = \Phi_{\Sigma}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n))$ with a Pearson product moment correlation matrix Σ .	$u_n = \Phi(A_{n1}\Phi^{-1}(\alpha_1) + \dots + A_{n(n-1)}\Phi^{-1}(\alpha_{n-1}) + A_{n(n)}\Phi^{-1}(\alpha_n))$ <p>where A_{ij} is the element of the Cholesky factorization that decomposes the covariance matrix Σ as $\Sigma = AA^T$ to give the lower triangular matrix $A = (A_{ij})_{i,j=1}^n$ and α_i is the percentile of the conditional distribution $X_i X_1, \dots, X_{i-1}$.</p>
t-copula	$C_T(u_1, \dots, u_n) = t_{\Sigma, \nu}(t_{\nu}^{-1}(u_1), \dots, t_{\nu}^{-1}(u_n))$ where t_{ν} is the univariate student's t distribution function, with ν degrees of freedom, and $t_{\Sigma, \nu}$ the multivariate distribution corresponding to t_{ν} .	$u_n = t_{\nu}(A_{n1}t_{\nu}^{-1}(\alpha_1) + \dots + A_{n(n-1)}t_{\nu}^{-1}(\alpha_{n-1}) + A_{n(n)}\sqrt{\frac{\nu + [t_{\nu}^{-1}(\alpha_1)]^2}{\nu + 1}}t_{\nu+1}^{-1}(\alpha_n))$ <p>where A_{ij} is the element of the Cholesky factorization that decomposes the covariance matrix Σ as $\Sigma = AA^T$ to give the lower triangular matrix $A = (A_{ij})_{i,j=1}^n$ and α_i is the percentile of the conditional distribution $X_i X_1, \dots, X_{i-1}$.</p>
Archimedean copula	$C_A(u_1, \dots, u_n) = \varphi^{[-1]}(\varphi(u_1) + \dots + \varphi(u_n))$ Where φ is the generator and $\varphi^{[-1]}$ is the inverse generator. The multivariate Archimedean Copula can also be defined as the n^{th} serial iterate of a two dimensional Archimedean copula (Jouini and Clemen 1996) as following: $C_A(u_1, \dots, u_n) = C_A(C_A(u_1, \dots, u_{n-1}), u_n)$	We can construct multivariate Archimedean copulas based on the interaction of bivariate Archimedean copulas since $C_A(u_1, \dots, u_n) = C_A(C_A(u_1, \dots, u_{n-1}), u_n)$. First, calculate $C_A(u_1, u_2)$; Second, draw u_3 from the conditional copula $C_A(u_3 C_A(u_1, u_2))$; More generally, draw u_n from $C_A(u_n C_A(u_1, \dots, u_{n-1}))$.

References

- Abbas, A. E. 2006. Entropy methods for joint distributions in decision analysis. *IEEE Trans. Engrg. Management* **53(1)** 146-159.
- Accioly, R., Chiyshi, F., 2004. Modeling dependence with copulas: a useful tool for field development decision process. *J. Pet. Sci. Eng.* **44** 83-91.
- Al-Harthy, M., S.Begg, R.B.Bratvold. 2007. Copulas: A new technique to model dependence in petroleum decision making. *J. Pet. Sci. Eng.* **57** 195-208
- Armstrong M., A. Galli, W. Bailey, B. Couet. 2004. Incorporating technical uncertainty in real option valuation of oil project. *J. Pet. Sci. Eng.* **44** 67-82
- Avramidis, A. N., N. Channouf, P. L'Ecuyer. 2009. Efficient Correlation Matching for Fitting Discrete Multivariate Distributions with Arbitrary Marginals and Normal-Copula Dependence. *INFORMS J. Comput.* **1(21)** 88-106.
- Brandão, L. E., J. S. Dyer, W. J. Hahn. 2005a. Using binomial decision trees to solve real-option valuation problems. *Decision Anal.* **2(2)** 69–88.
- Brandão, L. E., J. S. Dyer, W. J. Hahn. 2005b. Response to comments on Brandão et al. (2005). *Decision Anal.* **2(2)** 103–109.
- Bickel, E., J. E. Smith. 2006. Optimal Sequential Exploration: A Binary Learning Model. *Decision Anal.* **3(1)** 16–32
- Biller, B., B. L. Nelson. 2005. Fitting time-series input processes for simulation. *Oper. Res.* **53(3)** 549–559.
- Biller, B., 2009. Copula-Based Multivariate Input Models for Stochastic Simulation. *Oper. Res.* **57(4)** 878-892.
- Cario, M. C., B. L. Nelson. 1997. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix, Technical Report, Department of Industrial Engineering and Management Science, Northwestern University, Evanston, IL.
- Chen, H. 2001. Initialization for NORTA: Generation of Random Vectors with Specified Marginals and Correlations. *INFORMS J. Comput.* **13(4)** 312-331.
- Clayton, D.G., 1978. A model for association in bivariate life tables and its applications in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65** 141-151.
- Clemen, R. T., G. W. Fisher, R. L. Winkler. 2000. Assessing Dependence: Some Experimental Results. *Management Sci.* **46(8)** 1100–1115.
- Clemen, R. T., T. Reilly. 1999. Correlations and Copulas for Decision and Risk Analysis. *Management Sci.* **45(2)** 208–224.
- Clemen, R. T., T. Reilly., 2000. *Making Hard Decision with DecisionTools*, 2nd ed. Duxbury Press, Pacific Grove, CA.
- Cherubini,U., E. Luciano, W. Vecchiato.,2004. Copula Methods in Finance, John Wiley & Sons Ltd., New York.
- Conway, D. A. 1979. Multivariate Distributions with Specified Marginals. Technical Report#145, Department of Statistics, Stanford University.
- Copeland, T. E., V. Antikarov. 2001. Real Options: A Practitioner's Guide. Texere, New York.

- Darsow W. F., B, Nguyen, E. T. Olsen. 1992. Copulas and Markov processes. *Illinois J. Math.* **36**, 600-642.
- Embrechts, P., F. Lindskog, A. McNeil. 1999. *Modeling Dependence with Copulas and Applications to Risk Management*. Department of Mathematics, ETHZ CH-8092 Zurich, Switzerland.
- Frank, M.J. 1979. On the simultaneous associativity of $F(x,y)$ and $x+y-F(x,y)$. *Aequ. Math.* **19** 194-226.
- Ghosh, S., S. Henderson. 2002. Chessboard Distributions and Random Vectors with Specified Marginals and Covariance Matrix. *Oper. Res.* **50(5)** 820-834.
- Gumbel, E.J., 1960. Bivariate exponential distributions. *J. Am. Stat. Assoc.* **55** 698-707.
- Hult, H., Lindskog, F. 2002. Multivariate extremes, aggregation and dependence in elliptical distributions. *Adv. in Appl. Probab.* **34(3)** 587-608.
- Jouini, M., R. T. Clemen 1996. Copula Models for Aggregating Expert Opinions. *Oper. Res.* **44** 444-457.
- Keefer, D., S. E. Bodily 1983. Three-Point Approximations for Continuous Random Variables. *Management Sci.* **29** 595-609.
- Keefer, D. 1994. Certainty Equivalents for Three-Point Discrete-Distribution Approximations, *Management Sci.*, **40** 760-773.
- Kousky, C. and R. Cooke. 2009. The Unholy Trinity: Fat Tails, Tail Dependence, and Micro-Correlations. Resources For the Future (RFF) Discussion Paper 09-36-REV. Washington, D.C.
- Kotz S., J.R. van Dorp .2010. Generalized Diagonal Band Copulas with Two-Sided Generating Densities, *Decision Anal.*, **7(2)** 196-214.
- Kruskal, W. H. 1958. Ordinal Measures of Association. *J. Amer. Stat. Association.* **53(284)** 814-861.
- Li, S. T., J. L. Hammond. 1975. Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. *IEEE Trans. Syst., Man, and Cybernat.* **5** 557-561.
- Malevergne, Y., D. Sornette. 2005. Extreme Financial Risks: From Dependence to Risk Management. Springer.
- Nelsen, R.B. 1999. *An introduction to copulas*, Springer Berlin Heidelberg New York.
- Sklar, A. 1959. *Fonctions de répartition à n dimensions et leurs marges*, *Publ. Inst. Statist. Univ. Paris* **8**: 229-231.
- Smith, A. E., P. B. Ryan, J. S. Evans. 1992. The Effect of Neglecting Correlations When Propagating Uncertainty and Estimating the Population Distribution of Risk. *Risk Anal.* **12(4)** 467-474.
- Smith, J. E. 2005. Alternative approaches for solving real-options problems: (Comment on Brandão et al. 2005). *Decision Anal.* **2(2)** 89-102.
- Song, X., 2000. Multivariate Dispersion Models Generated from Gaussian Copula, *Scand. J. Statist.* **27** 305-320.
- Swain, J. J., S. Venkatraman, J. R. Wilson. 1988. Least-squares estimation of distribution functions in Johnson's translation system, *J. Stat. Comput. Simul.* **29** 271-297.
- Trivedi P. K., D. M. Zimmer. 2005. Copula Modeling: An Introduction for Practitioners. *Foundations and Trends® in Econometrics* **1(1)** 1-111.

Wang, T., J. S. Dyer, 2010. Valuing Multifactor Real Options Using an Implied Binomial Tree. *Decision Anal.* **7**(2) 185–195.

Wright, M.K., L. Stokes, J.S. Dyer. 1994. Reliability and Coherence of Causal, Diagnostic, and Joint Subjective Probabilities, *Decision Sci.* **25** (5) 691–709.

Yi, W. V. M. Bier. 1998. An Application of Copulas to Accident Precursor Analysis. *Management Sci.* **44** 257-270.