Correlations and Copulas for Decision and Risk Analysis

Robert T. Clemen • Terence Reilly
Fuqua School of Business, Duke University, Durham, North Carolina 27708
clemen@mail.duke.edu
Division of Math and Sciences, Babson College, Babson Park, Massachusetts 02157
reilly@babson.edu

The construction of a probabilistic model is a key step in most decision and risk analyses. Typically this is done by defining a joint distribution in terms of marginal and conditional distributions for the model's random variables. We describe an alternative approach that uses a copula to construct joint distributions and pairwise correlations to incorporate dependence among the variables. The approach is designed specifically to permit the use of an expert's subjective judgments of marginal distributions and correlations. The copula that underlies the multivariate normal distribution provides the basis for modeling dependence, but arbitrary marginals are allowed. We discuss how correlations can be assessed using techniques that are familiar to decision analysts, and we report the results of an empirical study of the accuracy of the assessment methods. The approach is demonstrated in the context of a simple example, including a study of the sensitivity of the results to the assessed correlations.
(*Measures of Dependence*; *Kendall's* τ; *Spearman's* ρ; *Copulas*; *Multivariate Normal Copula*; *Decision Analysis Process*)

1. Introduction

One of the central steps in decision and risk analysis is the construction of a model that portrays the uncertainty inherent in the situation. For example, such uncertainty could relate to risks associated with hazardous chemicals, uncertainty due to economic variables, or the stochastic nature of a manufacturing or service process. The conventional approach to modeling uncertainty is to specify a joint distribution of the random variables as a product of marginal and conditional distributions.

A disadvantage with the typical marginal-and-conditional approach is that the required number of probability assessments can grow exponentially with the number of variables. Analysts respond by searching diligently for conditional independence among variables to reduce the assessment burden. In this paper we discuss an alternative in which a joint distribution is constructed using a copula, requiring only marginal distributions and measures of dependence among the random variables. Using the copula that underlies the multivariate normal distribution, a complete copula-based joint distribution can be constructed using assessed rank-order correlations and marginal distributions, thereby reducing the number of required assessments and relaxing the need to search for conditional independence.

In the next section, we discuss the basics of constructing a copula-based probability model, showing how this approach can be implemented using the multivariate normal copula density. The copula approach relies fundamentally on the ability of experts to reliably assess correlations, which may appear to be a tall order. Some practical assessment techniques are available, however, and they are discussed in Section 3 along with the results of a small pilot study

0025-1909/99/4502/0208$05.00

investigating the techniques' accuracy. Section 4 demonstrates the copula procedure with an example, including a sensitivity analysis of the results to errors in assessed correlations. Section 5 concludes with a discussion of the benefits and limitations of the copula approach and our view of the technique's appropriate role as a complement to the conventional modeling approach.

2. Copula-Based Probability Models

Our starting point is the same as in a conventional decision analysis: We assume that the analyst has identified those uncertain variables for which a probabilistic model is required. For example, the analyst might perform a deterministic sensitivity analysis (Howard and Matheson 1983) or Reilly's (1998) dependent sensitivity analysis. With a set of random variables identified, the analyst proceeds by assessing marginal distributions and dependence measures necessary for constructing the copula-based joint distribution. Before we discuss specific copula models and assessment techniques, however, we review basic properties of copulas.

2.1. Copula Basics

The essence of the copula approach is that a joint distribution of random variables can be expressed as a function of the marginal distributions. To make this notion precise, we review two essential mathematical results. The first is:

Sklar's Theorem (1959). *Given a joint cumulative distribution function $F(x_1, \ldots, x_n)$ for random variables $X_1, \ldots, X_n$ with marginal cumulative distribution functions (CDFs) $F_1(x_1), \ldots, F_n(x_n)$, F can be written as a function of its marginals*:

$$F(x_1, \ldots, x_n) = C[F_1(x_1), \ldots, F_n(x_n)],$$

where $C(u_1, \ldots, u_n)$ is a joint distribution function with uniform marginals. Moreover, if each F_i is continuous, then C is unique, and if each F_i is discrete, then C is unique on $Ran(F_1) \times \cdots \times Ran(F_n)$, where $Ran(F_i)$ is the range of F_i.

The function C is called a *copula*. Sklar's Theorem is completely general: Any joint distribution can be written in copula form. Due to space limitations, we focus in this article on the mathematically tractable case where each F_i is continuous and differentiable.[1]

Given that F_i and C are differentiable, the joint density $f(x_1, \ldots, x_n)$ can be written as

$$f(x_1, \ldots, x_n) = f_1(x_1) \times \cdots \times f_n(x_n) c[F_1(x_1), \ldots, F_n(x_n)], \quad (1)$$

where $f_i(x_i)$ is the density corresponding to $F_i(x_i)$, and $c = \partial^n C / (\partial F_1 \cdots \partial F_n)$ is called the *copula density*. This is our second essential result, which states that, under appropriate conditions, the joint density can be written as a product of the marginal densities and the copula density. For example, if the X_i's are independent, then $c = 1$ and $f(x_1, \ldots, x_n) = f_1(x_1) \times \cdots \times f_n(x_n)$, the familiar formula for n independent random variables. From the representation in (1) it is clear that the copula density c encodes information about the dependence among the X_i's. For this reason c is sometimes called a *dependence function*.

To demonstrate the use of a copula to specify a joint distribution with specified marginals, suppose the analyst wishes to construct a bivariate distribution $H(x, y)$ with marginals $F(x)$ and $G(y)$. Specifically, let $F(x)$ be a beta distribution with parameters ($\alpha = 5$, $\beta = 5$), let $G(y)$ be a lognormal distribution with parameters ($\mu = 0$, $\sigma^2 = 1$), and let

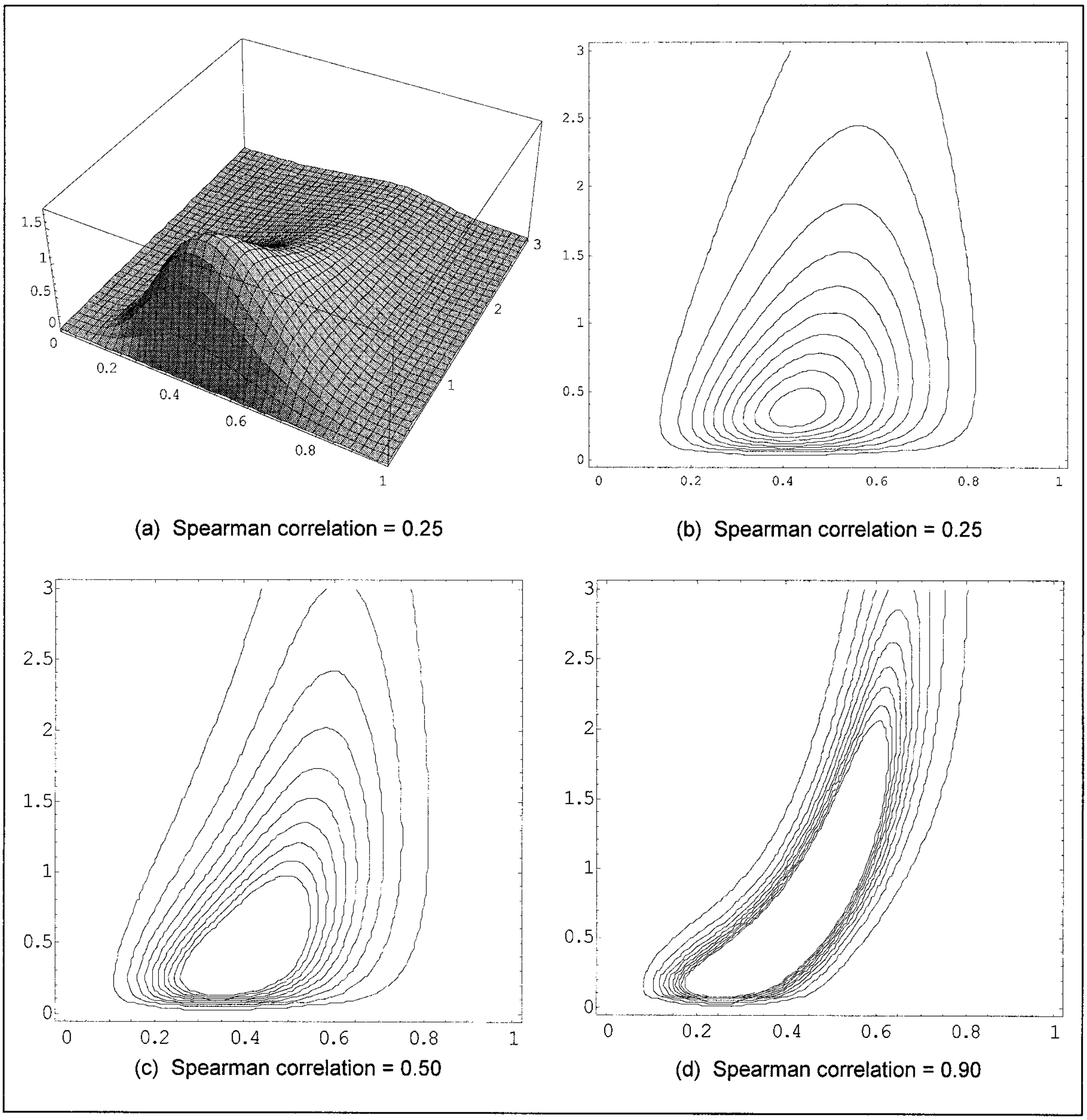
$$C_\delta(u, v) = \frac{-1}{\delta} \ln\left(1 + \frac{(e^{-\delta u} - 1)(e^{-\delta v} - 1)}{e^{-\delta} - 1}\right), \quad \delta \neq 0.$$

Then $C_\delta[F(x), G(y)]$ is a bivariate distribution of the requisite form $H(x, y)$. The copula $C_\delta(u, v)$ is a member of Frank's family (Frank 1979); the parameter δ determines the level of dependence between X and Y (Nelson 1986). As δ 3 0, $C_\delta[F(x), G(y)]$ approaches $F(x)G(y)$, implying independence. As δ 3 ∞, the correlation increases.

Figure 1 displays the joint density $h(x, y)$ for three different levels of correlation; in each case X and Y have the required beta and lognormal marginal densities.

[1] In the discrete case, the copula is unique on $Ran(F_1) \times \cdots \times Ran(F_n)$ rather than the unit hypercube because there is no unique definition for F_i^{-1}. With a specific definition of F_i^{-1}, the construction of a copula representation for a joint distribution of discrete random variables is straightforward. See the proof of Sklar's Theorem for further details and clarification.

Figure 1 **Bivariate Density with Beta (5, 5) and Lognormal (0, 1) Marginal Densities for Three Levels of Correlation**



Figures 1(a) and 1(b) show a three-dimensional plot and a contour plot of the bivariate density when X and Y have Spearman[2] rank correlation (ρ) of 0.25. Figures 1(c) and 1(d) illustrate via contour plots the effect of increasing the level of correlation to 0.50 and 0.90, respectively.

Using a copula as a basis for constructing a multivariate model is flexible because no restrictions are placed on the marginal distributions. Thus, we could just as easily have constructed a bivariate distribution with normal and gamma marginals or binomial and exponential marginals or any two subjectively as-

[2] Although Spearman's ρ was originally developed as a measure of association in a sample, it has a population analog that can be expressed in terms of a copula. The same is true for Kendall's τ. See Nelsen (1991).

sessed marginals. Such bivariate distributions are constructed by the substitution $u = F(x)$ and $v = G(y)$ as in the example above.

An excellent introduction to the theory of copulas is in Schweizer (1991). Tutorial articles include Genest and MacKay (1986) and Nelsen (1995). Dall'Aglio, Kotz, and Salinetti (1991) and Beneš and Štephán (1997) are collections of articles on copulas and related theory. A variety of applications involving copulas are available in the literature. For example, Jouini and Clemen (1996) use Archimedean copulas for aggregating expert judgments. Yi and Bier (1998) use copulas in an analysis of precursor events in a reliability model. Frees and Valdez (1998) and Frees, Carriere, and Valdez (1996) use copulas in actuarial modeling.

2.2. The Multivariate Normal Copula

Creating a copula-based probability model is accomplished by using the copula to "couple" the marginals into a joint distribution. Doing this requires two steps. First is modeling the marginal distributions in some way, which may require making a number of probability assessments and possibly fitting a member of a distribution family (normal, exponential, beta, etc.) to those assessments. Standard techniques from decision and risk analysis are available to accomplish these tasks (Morgan and Henrion 1990, Clemen 1996).

The second step is to create a copula to model the dependence among the random variables. Many concepts and measures of dependence or association are available. In this article, we consider only cases in which the relationships can be captured adequately by pairwise measures of dependence or association.[3,4] Two such measures are Spearman's ρ and Kendall's τ. For example, it is well known that if two random variables are positively (negatively) associated, then $\rho > 0$ ($\rho < 0$) and likewise for τ (Lehmann 1966, Barlow and Proschan 1975, Nelsen 1991). Moreover, these measures satisfy Schweizer and Wolff's (1981) *desiderata* for nonparametric measures of dependence and, for many families of copulas, can be used to index the family in terms of the level of dependence between the variables. Unlike the Pearson product-moment correlation, rank-order correlations such as ρ and τ do not depend on the marginal distributions. For now, we will assume that the expert has assessed a matrix $\mathbf{R}^*$ of dependence measures (either ρ or τ). Assessment of these two measures is discussed below in Section 3.

Several copula families are available that can incorporate the relationships defined by matrix $\mathbf{R}^*$. One such family is the copula that underlies the multivariate normal distribution. Like other copula families, the multivariate normal copula allows any marginal distribution for the X_i's. It is called the normal copula because it encodes dependence in precisely the same way that the multivariate normal distribution does using only pairwise correlations among the variables, but it does so for variables with arbitrary marginals. Moreover, the normal copula permits the use of any positive-definite correlation matrix. (The class of Archimedean copulas, for example, is limited to intraclass correlation matrices; see Jouini and Clemen (1996).) The flexibility and analytical tractability of the multivariate normal copula suggest that it is a promising way to represent dependence. Moreover, it has an illustrious history in probabilistic modeling. Although the notion of a copula *per se* is fairly recent, Edgeworth (1898) used the dependence structure of the bivariate normal to describe the joint distribution of bivariate data with clearly nonnormal marginals. Mardia (1970) also discusses a technique similar to what we use here. Kelly and Krzysztofowicz (1996a, b) use a closely related approach for selecting and combining expert forecasts.

To understand the multivariate normal copula, begin by recalling that the multivariate normal distribution typically is parameterized in terms of Pearson product-moment correlations. Thus, for each element of $\mathbf{R}^*$, calculate the corresponding product-moment correlation r_{ij} for the multivariate normal. If $\mathbf{R}^*$ is made up of τ_{ij}, the formula is $r_{ij} = \sin(\pi\tau_{ij}/2)$, and if the measure is ρ_{ij}, $r_{ij} = 2\sin(\pi\rho_{ij}/6)$ (Kruskal 1958). Construct matrix $\mathbf{R}$ with elements r_{ij}. Solving

[3] The concept of association is a relatively weak form of probabilistic dependence; see Barlow and Proschan (1975) for a definition.

[4] The approach we use permits only relationships that can be captured by pairwise rank correlations, which implies that ∪- or ∩-shaped relationships are not possible. If such a relationship is encountered, a suitable transformation of variables may render the problem tractable. Furthermore, we believe that limiting ourselves to such relationships is a natural starting point for representing expert knowledge.

Equation (1) for the copula density c_N in terms of the multivariate normal density $\phi^{(n)}(y_1, \ldots, y_n \mid \mathbf{R})$ gives:

$$c_N[\Phi(y_1), \ldots, \Phi(y_n) \mid \mathbf{R}^*] = \phi^{(n)}(y_1, \ldots, y_n \mid \mathbf{R}) / [\phi(y_1) \times \cdots \times \phi(y_n)], \quad (2)$$

where Φ and ϕ denote the univariate standard normal distribution and density, respectively. Substitution of the expressions for the normal densities and algebraic manipulation lead to:

$$c_N[\Phi(y_1), \ldots, \Phi(y_n) \mid \mathbf{R}^*] = \exp\{-\mathbf{y}^{\mathrm{T}}(\mathbf{R}^{-1} - \mathbf{I})\mathbf{y}/2\} / |\mathbf{R}|^{1/2}, \quad (3)$$

where $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$, and $\mathbf{I}$ is the $n \times n$ identity matrix.

We are now able to construct a multivariate density using c_N as a dependence function with arbitrary marginals $F_1(x_1), \ldots, F_n(x_n)$. Using the normal inverse transformation Φ^{-1}, define $Y_i = \Phi^{-1}[F_i(X_i)]$ for $i = 1, \cdots n$, and substitute these into (3) and (1) to obtain the desired joint density:

$$\begin{aligned} f(x_1, \ldots, x_n \mid \mathbf{R}^*) &= f_1(x_1) \times \cdots \times f_n(x_n) \\ &\quad \times \exp\{-\mathbf{y}^{\mathrm{T}}(\mathbf{R}^{-1} - \mathbf{I})\mathbf{y}/2\} / |\mathbf{R}|^{1/2} \\ &= f_1(x_1) \times \cdots \times f_n(x_n) \\ &\quad \times \exp\{-(\Phi^{-1}[F_1(x_1)], \ldots, \Phi^{-1}[F_n(x_n)]) \\ &\quad \times (\mathbf{R}^{-1} - \mathbf{I})(\Phi^{-1}[F_1(x_1)], \ldots, \\ &\qquad \Phi^{-1}[F_n(x_n)])^{\mathrm{T}}/2\} / |\mathbf{R}|^{1/2}. \end{aligned} \quad (4)$$

This joint density has the specified marginals and, because both ρ and τ are invariant under monotone 1-1 transformations of the original variables, the X_i's have the assessed rank-order correlations $\mathbf{R}^*$. Calculating the density for specific values $x_1, \ldots, x_n$ is relatively easy, requiring n inversions of the univariate standard normal distribution.

Conditional densities are also easily calculated using the multivariate normal copula model. Let $\mathbf{R}$ and $\mathbf{y}$ be partitioned as follows:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{n-1} & \mathbf{r} \\ \mathbf{r}^T & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = (\mathbf{y}_{n-1}, y_n),$$

where $\mathbf{y}_{n-1} = (y_1, \ldots, y_{n-1})^{\mathrm{T}}$, $\mathbf{R}_{n-1}$ is the $(n-1) \times (n-1)$ correlation matrix for $(Y_1, \ldots, Y_{n-1})$, and $\mathbf{r} = (r_{1n}, \ldots, r_{(n-1)n})^{\mathrm{T}}$. From (1), (2), and the definition of conditional probability, we have

$$f(x_n \mid x_1, \ldots, x_{n-1}, \mathbf{R}^*) = f_n(x_n) \times \frac{\phi^{(n)}(\Phi^{-1}[F_1(x_1)], \ldots, \Phi^{-1}[F_n(x_n)] \mid \mathbf{R})}{\phi(\Phi^{-1}[F_n(x_n)]) \times \phi^{(n-1)}(\Phi^{-1}[F_1(x_1)], \ldots, \Phi^{-1}[F_{n-1}(x_{n-1})] \mid \mathbf{R}_{n-1})},$$

which, upon substituting in the expressions for the normal densities and reducing, becomes

$$\begin{aligned} f(x_n \mid x_1, \ldots, x_{n-1}, \mathbf{R}^*) &= f_n(x_n) \exp\left\{-0.5\left[\frac{(\Phi^{-1}[F_n(x_n)] - \mathbf{r}^{\mathrm{T}}\mathbf{R}_{n-1}^{-1}\mathbf{y}_{n-1})^2}{(1 - \mathbf{r}^{\mathrm{T}}\mathbf{R}_{n-1}^{-1}\mathbf{r})}\right.\right. \\ &\quad \left.\left. - (\Phi^{-1}[F_n(x_n)])^2\right]\right\}(1 - \mathbf{r}^{\mathrm{T}}\mathbf{R}_{n-1}^{-1}\mathbf{r})^{-1/2}. \end{aligned} \quad (5)$$

With the joint density specified, expected values, expected utilities, and risk profiles can be calculated directly from the copula model. Value-of-information analysis is possible and will typically require the calculation of conditional distributions from the copula in a manner similar to (5) above. As usual in the decision-analysis process, modeling and analysis may iterate until clarity of action is obtained. The example in Section 4 demonstrates both discrete-approximation and simulation approaches for performing some of the calculations that might be required in an analysis.

3. Assessing Correlations

Using the multivariate normal copula requires the assessment of correlations. Although useful methods for assessing probabilities are well known, analysts typically do not try to assess moments of distributions (Morgan and Henrion 1990). The assessment of cross moments would appear to be an even more difficult problem. Although such judgments are inherently difficult, three general approaches are available to help an expert think about relationships among random variables. In this section we describe three assessment approaches, report the results of an initial empirical study comparing the methods' accuracy, and discuss additional assessment concerns.

3.1. Correlation Assessment Methods

3.1.1. Statistical Approaches. These techniques rely on an expert's familiarity with statistical concepts related to correlation. For example, an expert might make a judgment regarding the "percentage of variance explained" (R^2) that would result from regressing one variable on another. Another possibility is to have the expert view several scatterplots representing different levels of correlation and select one that is consistent with the expert's belief about the strength of the relationship between the variables. This approach is currently used informally in *Crystal Ball*, a popular risk-analysis add-in for Microsoft® Excel™. Experts who have substantial training in statistical data analysis may be able to think of bivariate relationships easily in these terms. Gokhale and Press (1982), for example, show that individuals with statistical training are capable of viewing a scatterplot of bivariate data and making a reasonably accurate assessment of the sample correlation of those data.

3.1.2. Probability of Concordance. The second approach to correlation assessment is consistent with decision analysis elicitation techniques: Assess conditional or joint probabilities and relate those to the required measure of dependence. Here we use the concordance probability P_C, following Gokhale and Press (1982). For a bivariate population (X, Y), we define P_C by considering two independent draws (x_1, y_1) and (x_2, y_2):

$$\begin{aligned} P_C &= P[(x_1 \le x_2 \quad \text{and} \quad y_1 \le y_2) \text{ or} \\ &\qquad (x_1 > x_2 \quad \text{and} \quad y_1 > y_2)] \\ &= P(x_1 \le x_2 \mid y_1 \le y_2). \end{aligned}$$

P_C can be related to Kendall's τ:

$$\tau = 2P_C - 1.$$

The primary difficulty with using P_C and τ is the nature of the joint or conditional event for which the probability must be assessed. As long as a natural interpretation of the event in frequency terms exists, this is a minor problem. For example, judging the relationship between height and weight in a population by assessing P_C would require the expert to answer a question like the following:

> "Two individuals (A and B) are chosen randomly from a population of adult males. Given that A weighs less than B, what is the probability that A is also shorter than B?"

Consider, however, what the assessment question would look like if it concerned the relationship between the size (S) of a population of organisms (say, an endangered species of tree frogs) and a temperature index T for the frogs' environment. In such a case, there may be only one population of such frogs and no obvious frequency interpretation to relate S and T. The expert would have to answer a difficult question like the following:

> "Imagine a hypothetical situation in which there are two separate and independent populations of tree frogs, each one with its own population size and temperature measurement. Call these two situations A and B, with corresponding population size and temperature pairs (s_A, t_A) and (s_B, t_B). Given that t_A is less than t_B in this hypothetical situation, what is the probability that s_A is also less than s_B?"

As long as the concordance probability is assessed for a situation in which a frequency or cross-sectional interpretation is reasonable (e.g., relationships among security returns or among variables related to cancer risk), the less complex form of the assessment question may make the assessor's job easier. Moreover, frequency framing suggests that the expert would be less susceptible to cognitive biases (Gigerenzer 1991, Gigerenzer et al. 1991). The method may be less useful for applications involving one-time events, such as the extinction of a species or global warming.

3.1.3. Conditional Fractile Estimates. The third method requires conditional estimates and uses these to derive Spearman's ρ. Given random variables X and Y with corresponding distribution functions $F(x)$ and $G(y)$, the standard nonparametric regression representation is:

$$E[F(X) \mid y] = \rho_{XY}[G(y) - 0.5] + 0.5, \qquad (6)$$

where ρ_{XY} is the Spearman correlation between X and Y. Equation (6) suggests that the expert be asked questions like the following:

> "Suppose an individual is randomly chosen from a population of adult males. Given that his height is found to be 180 cm, the 70th percentile of the distribution of heights for this population, what would you estimate for the percentile at which his weight falls?"

If the marginal distributions have been assessed, as would have been done in a typical analysis, then the percentile estimate can be related directly to the marginal distribution; for example, the expert might respond that the conditional estimate of the weight would be the 60^{th} percentile of the marginal distribution of weights, corresponding to 85 kg. Given $G(180) = 0.70$ and $E[F(X) \mid 180] = 0.60 = F(85)$, the analyst can solve to find that $\rho_{XY} = 0.50$. The analyst might have the expert make several conditional estimates and use a least-squares approach to estimate ρ_{XY}. This method is closely related to a technique called *predictive assessment* originally developed by Bayesian statisticians for assessing a prior distribution for the parameters in a linear model (Winkler et al. 1978, Kadane et al. 1980).

3.2. Empirical Results

To study the ability of individuals to make correlation assessments, we performed a small pilot study. (More elaborate studies are underway at Duke University by Clemen and colleagues.) In November 1997, 58 Babson College students (30 MBA students and 28 business undergraduates) responded to three questions (probability of concordance, conditional fractile, and direct estimate of the correlation) for three different pairs of financial variables: monthly returns for the Standard and Poor's 500 and the Dow Jones Industrial Average; monthly returns for shares of Chrysler Corporation and an Automotive Index compiled by Compustat; and monthly returns for shares of Chrysler and Eli Lilly and Company. We will refer to these three pairs of variables as SD, CA, and CE, respectively, for which the actual correlations are 0.95, 0.74, and 0.17 as calculated from the most recent 20 years of Compustat data. All of the students had previously been exposed to the concept of correlation. The average number of years of experience in the financial-securities industry was 0.80 for graduates and 0.43 for undergraduates. Only seven students in each group reported such experience.

Each subject began by answering sample questions about the relationship between height and weight for male MBA students. For each of the sample questions, the experimenter read a short statement explaining how to think about the assessment question. Once the sample questions were completed, each subject answered the same types of questions for the three pairs of financial variables. For each pair of variables, each subject viewed histograms of the marginal distributions as well as the means, standard deviations, and deciles of the marginals. No scatterplot was provided (as was done in Gokhale and Press (1982)) because we were interested in subjects' ability to convert their knowledge into accurate dependence assessments. To counter order effects, a completely randomized design was used. Complete details regarding the design and questionnaire are available on request from the authors.

Differences between the responses of graduates and undergraduates were not significant. Thus, we pooled the two groups for the subsequent analysis. To compare the assessment methods, we began by converting conditional fractile and concordance probability responses to equivalent Spearman correlations using formulas from Section 2. The equivalent correlations then were compared to the actual correlations. For each subject and each assessment, we calculated the absolute error as the absolute value of the difference between the actual and equivalent assessed correlation. Various summary statistics are presented in Table 1.

First, we note that all of the assessment methods are meaningful to the subjects; average assessed correlations are highest for SD, next highest for CA, and lowest for CE. The subjects tended to underestimate the correlations for the two most highly correlated pairs, SD and CA.

The average absolute errors indicate the level of accuracy of the correlation assessments. Some differences are apparent for the different pairs, but the level of accuracy is remarkably consistent across the three assessment methods. Averaging over all three variable pairs, the average absolute error falls consistently near 0.25.

The subjects also rated the difficulty of the assessment methods on a 7-point Likert scale. Direct assessment of the correlation was viewed as somewhat easier than the two indirect methods. Given the students' prior exposure to correlation, this result does not appear surprising.

3.3. Additional Correlation-Assessment Concerns

All three assessment methods require some training. For direct assessment of the correlation, the expert

Table 1 Summary Statistics for Correlation-Assessment Study. The Estimated Correlations are Based on 20 Years of Compustat Data

	Assessment Method			
	Direct Assessment	Concordance Probability	Conditional Fractile	Estimated Correlation
AVERAGE ASSESSED CORRELATIONS				
S&P 500 and Dow Jones	0.81	0.78	0.77	0.95
Chrysler and Auto Index	0.62	0.61	0.66	0.74
Chrysler and Eli Lilly	0.25	0.09	0.21	0.17
AVERAGE ABSOLUTE ERRORS				
S&P 500 and Dow Jones (SD)	0.19	0.18	0.22	
Chrysler and Auto Index (CA)	0.22	0.26	0.20	
Chrysler and Eli Lilly (CE)	0.32	0.31	0.31	
All Variables	0.24	0.25	0.24	
AVERAGE DIFFICULTY RATING	3.00	3.76	3.78	

should be trained in the statistical concept of correlation. For the concordance probability, training may include learning about probability assessment as well as clarification of the concordance event. For the conditional fractile estimate, the expert must understand fractiles. In addition, the expert must also understand the notion of regression toward the mean. It is evident from (6) that $|E[F(X) \mid y] - 0.5| \leq |G(y) - 0.5|$. Thus, in our example where $G(180) = 0.70$, $E[F(X) \mid Y = 180]$ must fall between 0.30 and 0.70.

On a technical issue, **R** must be positive definite. If correlations are assessed individually without consideration of the overall joint distribution, the resulting **R** may be nonpositive definite, requiring reassessment. Even a poorly conditioned **R** (analogous to multicollinearity in regression) may lead to some very counterintuitive results; for example, conditional distributions and risk profiles may appear deceptively narrow, reflecting the mathematics of the specific model, although such a result may be inconsistent with the expert's intuition.

Meeuwissen and Cooke (1994) and Cooke (1995) describe "dependence trees" along with an entropy-maximization approach for generating the correlation matrix. By assessing dependence measures hierarchically via a dependence tree, fewer assessments are required, but the nature of the dependence structure that can be modeled is somewhat constrained. Entropy maximization ensures that the resulting correlation matrix is positive definite.

Each of the three methods for correlation assessment described above has advantages and limitations. In view of the arguments as well as our (limited) empirical results, we make no claim for the superiority of one method over another. For now, we recommend that analysts and experts use a combination of approaches. For example, an initial assessment of Spearman's ρ can be used to generate a sample scatterplot based on simulated values. Likewise, ρ can be used to derive a corresponding P_C, and the expert can be asked if the implied P_C adequately reflects his or her reasoning.

4. An Example: Eagle Airlines

4.1. Initial Model Specification and Deterministic Sensitivity Analysis

Clemen (1996) describes the hypothetical decision faced by Dick Carothers, owner of the fledgling Eagle Airlines. Carothers is considering purchasing a used aircraft. His decision criterion is whether the airplane will generate more profit than a money-market alternative investment. Reilly (1998) modifies the model slightly in his sensitivity-analysis example, and we will use Reilly's model here. The influence diagram in Figure 2 portrays the initial model.

The first step is to conduct a sensitivity analysis to identify the critical variables, either with the standard one-way sensitivity analysis (Howard and Matheson 1983, Clemen 1996) or dependent sensitivity analysis as described by Reilly (1998). As Reilly demonstrates, in this case the two sensitivity-

Figure 2 **The Initial Influence Diagram for Eagle Airlines**

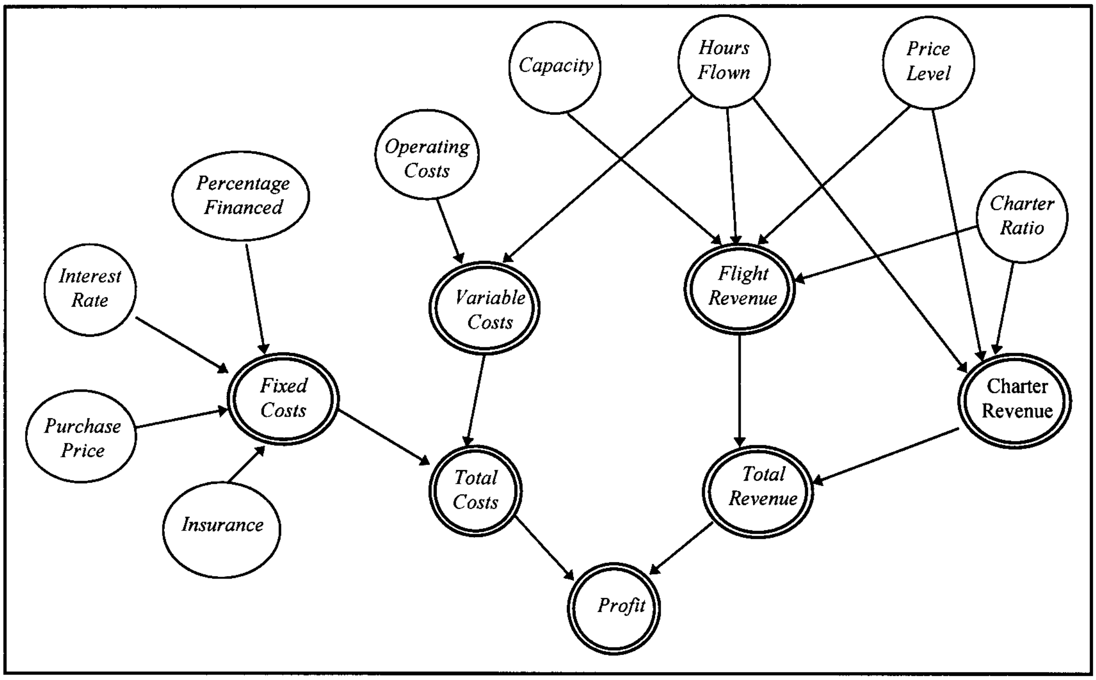


Table 2 **Fractiles and Spearman Correlations for Four Critical Variables in Eagle Airlines**

				Spearman Correlations		
Variable (X) Fractile:	Low 0.10	Base 0.50	High 0.90	*Price Level*	*Hours Flown*	*Capacity*
Price Level (P)	\$ 95	\$100	\$ 108			
Hours Flown (H)	500	800	1000	−0.50		
Capacity (C)	40%	50%	60%	−0.25	0.50	
Operating Cost per Hour (O)	\$230	\$245	\$ 260	0	0	0.25

analysis approaches produce slightly different insights but eventually identify the same set of critical variables: *Price*, *Hours Flown*, *Capacity*, and *Operating Cost*, which we will denote by P, H, C, and O, respectively. Information on these variables, including 0.10, 0.50, and 0.90 fractiles, along with hypothetical Spearman correlations, are shown in Table 2.

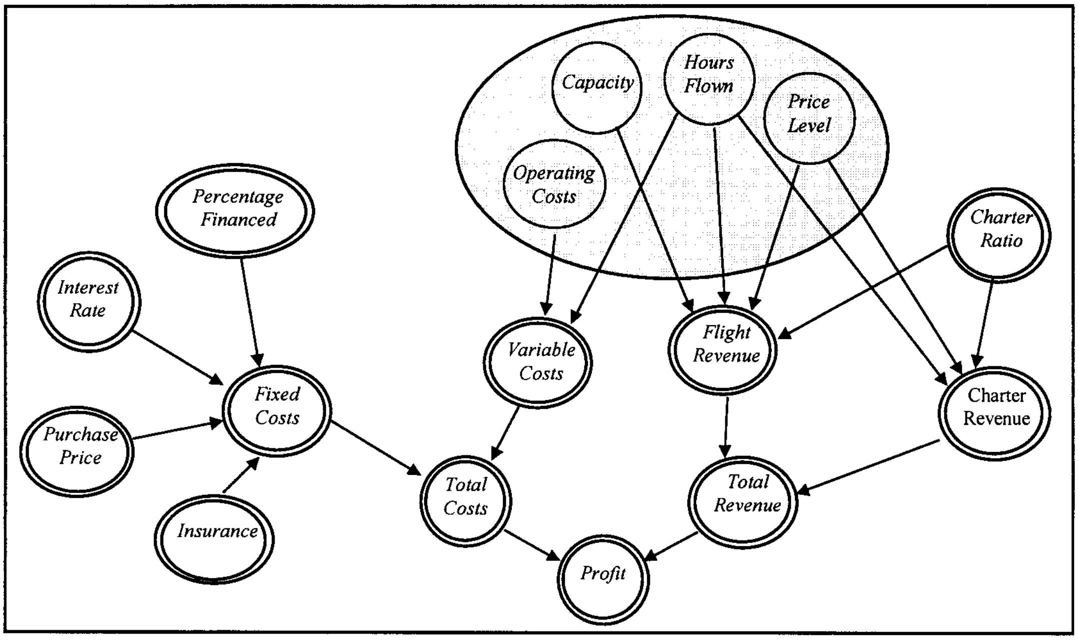
4.2. A Copula-Based Joint Density

To create a copula-based joint density, we first model the marginal densities for the four variables as indicated in Table 3. The modeled densities have approximately the same 0.10, 0.50, and 0.90 fractiles as shown in Table 2. Denote the marginal beta density and cumulative distribution for P as $f_\beta(p)$ and $F_\beta(p)$,

Table 3 **Marginal Distributions for 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	$\mu = 245$, $\sigma = 11.72$	$(-\infty, +\infty)$

Figure 3 Representing a Copula-Based Probability Model in an Influence Diagram



respectively, and similarly for H and C.[5] Likewise, let $f_N(o)$ and $F_N(o)$ denote the marginal density and cumulative distribution for the normally distributed O.

The second step in creating the joint density is to specify the copula c_N. This is a matter of using the assessed Spearman correlations to calculate $\mathbf{R}$ and substituting $\mathbf{R}$ and the marginal densities and distributions into (4). For Eagle Airlines, we substituted the marginal densities from Table 3, the corresponding distributions (F_i), and $\mathbf{R}$ into (4). Let $y_p = \Phi^{-1}[F_\beta(p)]$, $y_h = \Phi^{-1}[F_\beta(h)]$, $y_c = \Phi^{-1}[F_\beta(c)]$, $y_o = \Phi^{-1}[F_N(o)]$, and let $\mathbf{y} = (y_p, y_h, y_c, y_o)$. We have:

$$f(p, h, c, o) = f_\beta(p) f_\beta(h) f_\beta(c) f_N(o) \times \exp\{-\mathbf{y}^{\mathrm{T}}(\mathbf{R}^{-1} - \mathbf{I})\mathbf{y}/2\}/0.486, \quad (7)$$

where

$$\mathbf{R}^{-1} - \mathbf{I} = \begin{bmatrix} 0.366 & 0.715 & -0.014 & 0.004 \\ 0.715 & 0.777 & -0.787 & 0.205 \\ -0.014 & -0.787 & 0.506 & -0.393 \\ 0.004 & 0.205 & -0.393 & 0.103 \end{bmatrix}.$$

Influence diagrams can be very useful for communicating with decision makers about an analytical model, but at present no convention exists for representing a copula-based joint distribution in an influence diagram. Any convention adopted must accommodate the presence of predecessors and successors that are external to the copula portion of the model. At the same time, directed arcs among the variables that are related by the copula are inappropriate. We suggest placing the copula-related variables near each other and within a shaded region as in Figure 3. Each variable in the joint distribution will require assessment of its marginal distribution only, and dependence measures among the variables must be assessed to specify the copula. Explanation of the model might include a statement to the effect that the variables in the copula model are considered together, and relationships among them (correlations) are explicitly accounted for in the model.

[5] The scaled beta density $f_\beta(x)$ is created by specifying a closed interval $[x_0, x_1]$ for the support of the random variable X and then calculating the density as

$$f_\beta(x \mid \alpha, \beta, x_0, x_1) = \{[(x - x_0)/(x_1 - x_0)]^{\alpha-1}[(x_1 - x)/(x_1 - x_0)]^{\beta-1}\}/[\mathrm{B}(\alpha, \beta)(x_1 - x_0)],$$

where $\mathrm{B}(\alpha, \beta)$ is the beta function.

4.3. A Discrete Approximation

Analysis of the four-dimensional continuous density in (7) requires numerical integration or Monte Carlo simulation. In the spirit of current practice in decision analysis and to demonstrate the flexibility of the copula approach, we show in this section how a discrete approximation can be created. The approach we take is to calculate conditional densities using (5) and use these to create an event tree based on the extended Pearson-Tukey discrete approximation procedure (Keefer and Bodily 1983). The robust performance of this method is documented by Keefer and Bodily as well as Keefer (1994) and Runde (1997).

Using (5), it is straightforward to derive the conditional densities $f(h \mid p)$, $f(c \mid h, p)$, and $f(o \mid c, h, p)$:

$$\begin{aligned} f(h \mid p) = f_\beta(h) \exp\{-0.5[(\Phi^{-1}[F_\beta(h)] \\ + 0.518\Phi^{-1}[F_\beta(p)])^2/0.732 \\ - (\Phi^{-1}[F_\beta(h)])^2]\}/0.732^{1/2}, \end{aligned}$$

$$\begin{aligned} f(c \mid p, h) = f_\beta(c) \exp\{-0.5[(\Phi^{-1}[F_\beta(c)] \\ - 0.009\Phi^{-1}[F_\beta(p)] \\ - 0.523\Phi^{-1}[F_\beta(h)])^2/0.732 \\ - (\Phi^{-1}[F_\beta(c)])^2]\}/0.732^{1/2}, \end{aligned}$$

$$\begin{aligned} f(o \mid p, h, c) = f_N(o) \exp\{-0.5[(\Phi^{-1}[F_N(o)] \\ + 0.003\Phi^{-1}[F_\beta(p)] \\ + 0.186\Phi^{-1}[F_\beta(h)] \\ - 0.357\Phi^{-1}[F_\beta(c)])^2/0.907 \\ - (\Phi^{-1}[F_N(o)])^2]\}/0.907^{1/2}. \end{aligned}$$

As indicated, each conditional density is approximated using the extended Pearson-Tukey method. Figure 4 shows a portion of the resulting event tree. Throughout the tree, the values displayed on the branches are the levels of the variables at the 0.05, 0.50, and 0.95 fractiles of the corresponding marginal or conditional density. Probabilities are not shown on the tree because the same trio of probabilities (0.185, 0.63, and 0.185, *per* the extended Pearson-Tukey formula) are applied at each chance node. The complete event tree is available on request from the authors.

The bivariate relationships among the variables reveal themselves readily in Figure 4. For example, as *Price* increases, *Hours Flown* tends to decrease, reflecting the negative correlation between these two variables, and as *Hours Flown* increases, its positive correlation with *Capacity* tends to lead to an increase in the latter.

4.4. Analysis

We can analyze the Eagle Airlines problem by using the multivariate normal copula model (7) directly in a Monte Carlo simulation as described in the *UNICORN* User's Manual (Cooke 1995). First generate a vector (y_p, y_h, y_c, y_o) from a multivariate-normal process with correlation matrix **R**. The standard normal distribution function $\Phi(y_i)$ is calculated for each of the four y variables. Finally, use the inverse marginal distribution functions to calculate $(p, h, c, o) = (F_p^{-1}[\Phi(y_p)], F_h^{-1}[\Phi(y_h)], F_c^{-1}[\Phi(y_c)], F_o^{-1}[\Phi(y_o)])$. This vector of variates comes from a process that has the specified marginal distributions as well as the required rank correlations.

Working in Microsoft® Excel™ and Crystal Ball®, we used the simulation procedure described above to estimate the expected profit, standard deviation, and risk profile for Eagle Airlines. We also ran the simulation under the assumption that the variables are independent. Both simulations were run for 10,000 trials. We also calculated the expected profit, standard deviation, and risk profile using the discrete approximation from Section 4.3.

The results are displayed in Table 4 and Figure 5. If we take the copula-based simulation as the benchmark, some observations can be made. First, although the independence simulation accurately estimates the expected profit—as one would anticipate—it produces a larger standard deviation (a difference of more than \$3400). The discrete approximation, which does incorporate dependence, suffers from jumps in the CDF near the median (as one would expect from a discrete approximation with the bulk of the probability assigned to the median). Nevertheless, the discrete approximation provides a reasonable estimate of the expected profit (with a difference of less than \$200), the standard deviation (a difference of less than \$80), and a very reasonable fit to the tails of the copula-based risk profile.

Given the assessment accuracy results reported in Section 3, an important question is how sensitive these

Figure 4 Event Tree from Multivariate Normal Copula

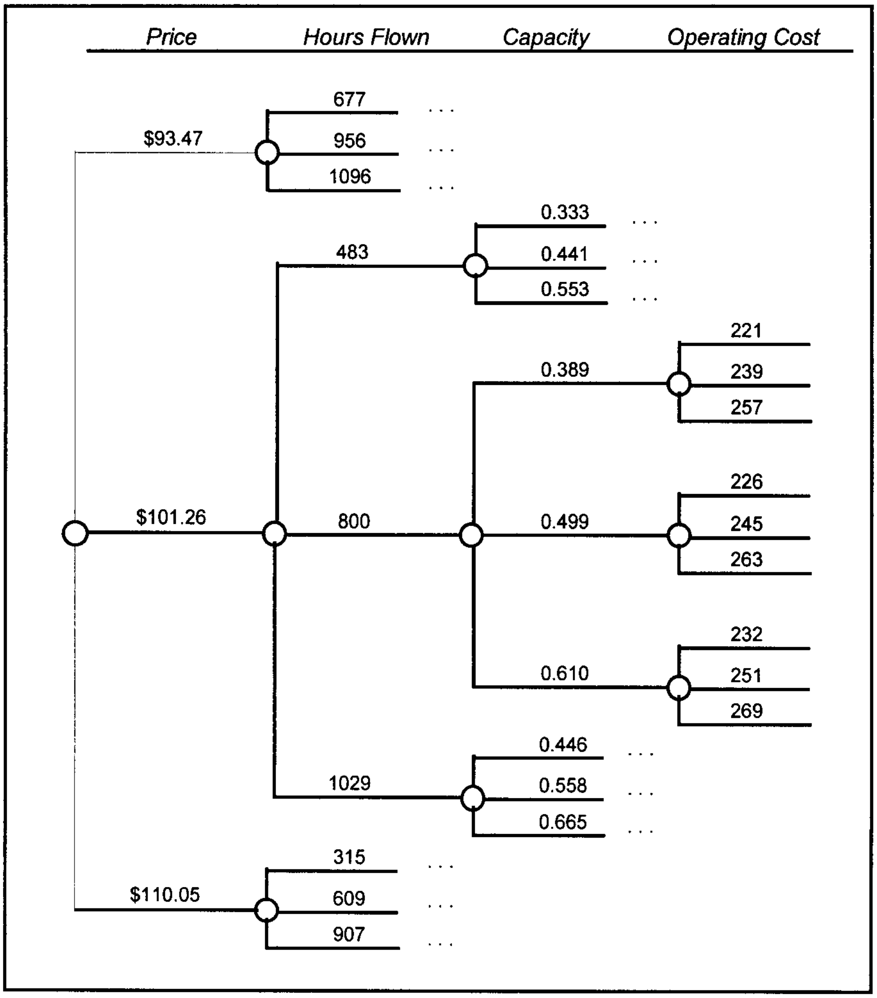


Table 4 Expected Profit and Standard Deviation for Eagle Airlines Calculated from Three Models

	Expected Value	Standard Deviation
Copula simulation (10,000 trials)	$12,417	$20,206
Independent simulation (10,000 trials)	$12,426	$23,628
Discrete approximation	$12,606	$20,281

results are to the specified correlations. To investigate the influence of the correlation matrix on the decision model, we used different matrices as inputs into the copula model, calculating means and standard deviations of the resulting risk profiles.

We performed two sensitivity analyses. The first explores the range of possible output values for any level of dependence, i.e., for any positive-definite correlation matrix. Our objective was to find four matrices, one each that produce the greatest and least average profit, and one each that produce the greatest and least standard deviation of profit. Our results indicate that the average profit can range from a low of $6552 to a high of $22,049, and that the standard deviation can range from $9493 to $45,215.[6]

[6] The problem of precisely identifying the correlation matrices that generate extreme means and standard deviations is a complex optimization problem. We approached the problem heuristically as follows: First, we randomly generated 10,000 4 $\times$ 4 correlation

In the second sensitivity analysis, we perturbed the nonzero correlations in **R** by ±0.25, approximately the average MAD from our pilot study. (We left the two zero elements fixed, reflecting the analyst's modeling process; setting these values to zero is primarily a statement regarding the structure of the relationships among the variables, whereas the nonzero elements require error-prone assessments as discussed in Section 3.) We considered all possible combinations in which each of the four correlations could take on three values: its base value r_{ij}, $r_{ij} + 0.25$, or $r_{ij} - 0.25$. This yielded 81 different matrices. Of these, five were non-positive-definite and were eliminated. For each of the remaining 76 matrices, we ran the simulation for 3500 trials (convergence of means to within 1.5% occurs at about 2000 trials), collecting the mean and standard deviation of the resulting risk profile. Table 5 shows percentiles of the means and standard deviations for these 76 simulation runs. For example, 50% of the 76 means were less than or equal to \$12,610.

Figure 6 combines the results of the two sensitivity analyses graphically. It shows clearly that all of the means and standard deviations from the perturbed correlation matrices fall well within the extremes found in the first sensitivity analysis, and in many cases varying the correlations by 0.25 does not lead to vastly different output results. For example, 50% of the cases (those between the first and third quartiles) have means that fall within 7% of the base case and standard deviations within 13% of the base case. Likewise, 80% of the cases fall within 13% (22%) of the mean (standard deviation) of the base case. Thus, although individuals may be able to assess correlations only with limited precision (as might be argued with probability assessment as well), for the example studied here the typical degree of assessment error may not be crucial. Of course, whether a given level of assessment accuracy is adequate depends on the specific decision problem at hand.

Others have also investigated the impact of correlations in risk-analysis models. For example, a variety of articles in the risk analysis literature have advocated the importance of including correlations in Monte Carlo simulation models rather than behaving as if the variables are independent (e.g., Apostolakis and Kaplan 1981, Burmaster and Anderson 1994, Kraan and Cooke 1997). As in our initial Eagle Airlines analysis, these articles show that an assumption of independence can be misleading. Smith, Ryan, and Evans (1992) show conditions under which correlations can be ignored, including situations where the correlations are weak, where there is relatively little uncertainty about the variables in question, or where the variables have relatively little influence on the outcome measure. Lacke's (1998) work is the most closely related to our sensitivity analysis. He develops a colorectal cancer risk model in which expert distributions are combined using the copula model from Jouini and Clemen (1996). Lacke studies several different cancer-screening policies. Varying the correlations used in the copula aggregation model from 0.90 to 0.50 led to no changes in the rank order of the top four policies.

5. Discussion and Conclusion

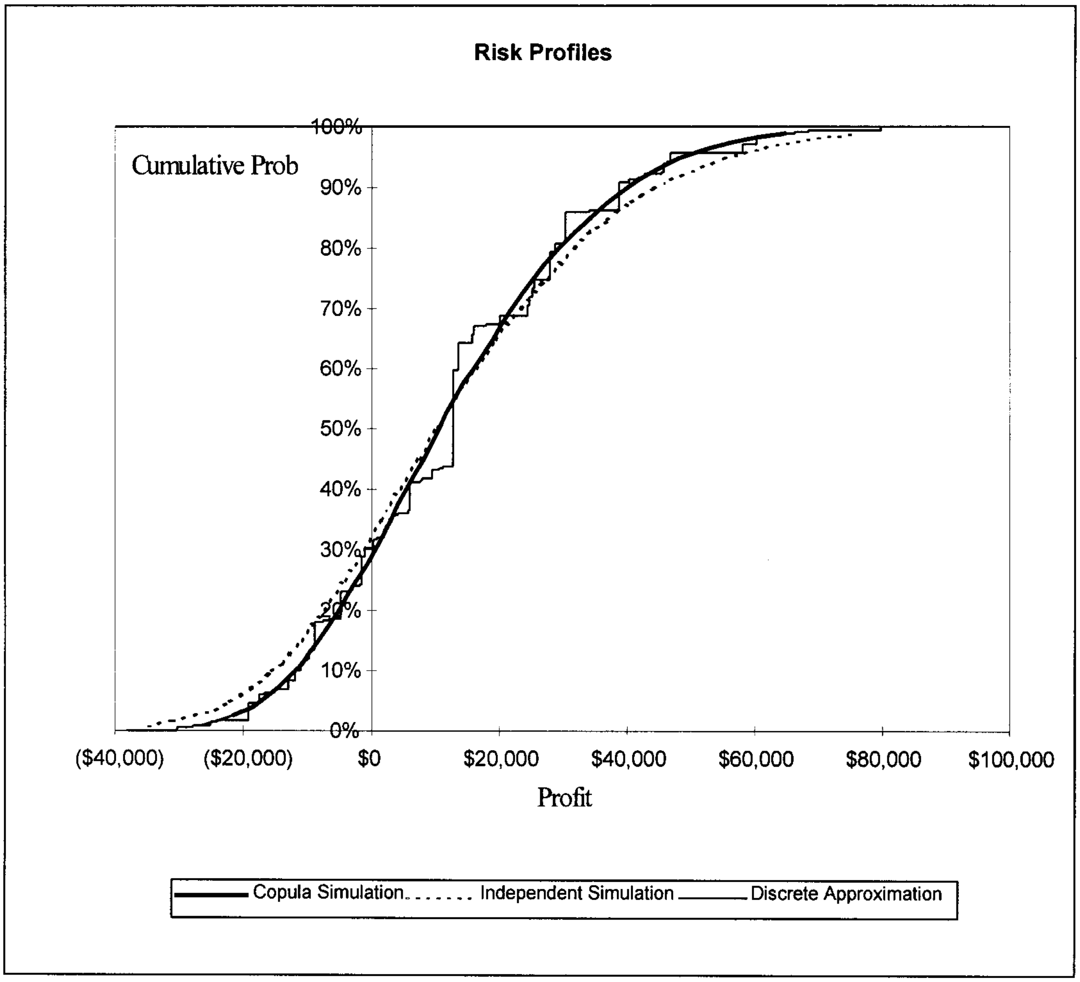
We have argued that decision and risk analysis can benefit from the use of correlations and copulas in the construction of probabilistic models, and we showed explicitly how to realize these benefits via practical correlation-assessment methods and the multivariate normal copula. This approach complements the set of tools available to the analyst, potentially streamlining the construction and analysis of probabilistic models and reducing the assessment burden. In this final section we discuss several of the issues surrounding the use of correlations and copulas.

With regard to the modeling process itself, there are

matrices and discarded those which were not positive definite. Our (unproven) conjecture is that the extreme results we sought would be produced by matrices that lie near the boundary of positive-definiteness, and so we restricted further investigation to those having determinants less than 0.01, of which there were 114 in our sample of 10,000. Using these correlation matrices as inputs in the simulation generated the required summary statistics. Once the four matrices were identified, slight perturbations were applied manually, based on reasoning about the algebraic structure of the value function, in an effort to extend the boundary further.

Our intent was to use a reasonably efficient heuristic procedure to find cases we believed to be close to the boundary. Thus, although our results may not lie precisely on the boundary, the true boundaries must be at least as extreme as the ones we found. Further details, including the matrices that generated our boundary results, are available on request from the authors.

Figure 5 **Risk Profiles for Eagle Airlines from Three Models**



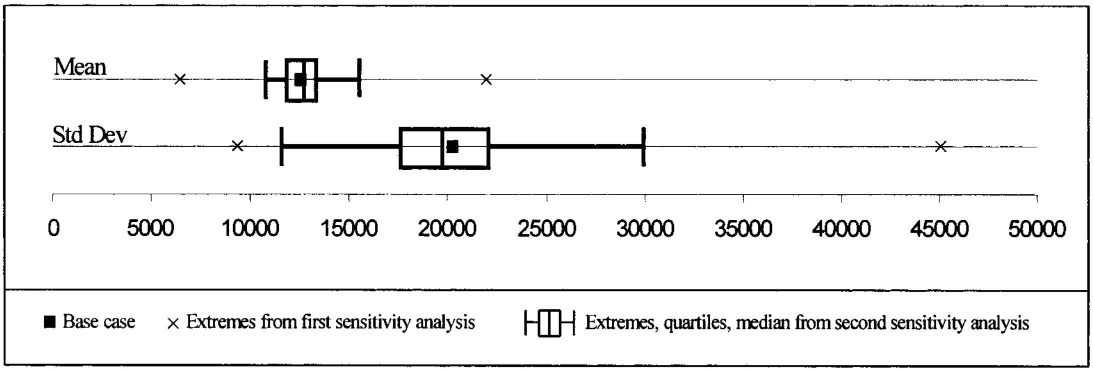
many ways to construct and use copula-based joint densities that were not mentioned in Sections 2 and 3. For example, we focused on the multivariate normal copula, but others may also be useful, such as the maximum-entropy copulas described by Meeuwissen (1993) and MacKenzie (1994). In some cases, decision analysts discretize density functions; we used the extended Pearson-Tukey discrete-approximation method, but other methods may also be used. Modeling the entire joint distribution raises the question of how best to discretize a full joint distribution: How does one select appropriate representative points, and what probabilities should be applied? This question amounts to choosing representative scenarios in a complex multivariate space. (See DeVuyst, Preckel, and Liu (1999).)

The calculation of conditional distributions from the copula-based joint distribution is straightforward. One implication of this is that the search for conditional independence becomes less critical; it yields no real savings either in assessment or computational complexity. The ease of calculating conditional distributions will be especially useful for more complex models or analyses. For example, in a value-of-information analysis, it would be necessary to separate out one or more variables from the copula, conditioning decisions and the remaining copula variables on the information variable. For inference

Table 5 Percentiles of Means and Standard Deviations of Risk Profiles Resulting from Simulations with Assessed Correlations Perturbed by ±0.25 in Various Combinations as Described in the Text. The "Base Case" Column Contains the Mean and Standard Deviation from the Original Simulation Using the Assessed R.

	Base Case	Percentile: 0%	10%	25%	50%	75%	90%	100%
Mean	$12,417	$10,677	$11,409	$11,760	$12,610	$13,338	$14,034	$15,388
Std Dev	$20,206	$11,454	$15,827	$17,486	$19,645	$21,996	$24,194	$29,887

Figure 6 Results from Two Sensitivity Analyses. The Box-and-Whisker Plots Show the Distribution of Means and Standard Deviations of Profit in the Eagle Airlines Problem Resulting from Perturbing the Correlations by ±0.25. As Shown in this Graph, Assessment Errors of this Magnitude Typically do not Have a Large Impact on the Results of the Analysis



applications, the ease of calculating the conditional distributions should make it straightforward to propagate information through a network.

Although this article has focused on the case when the marginals are continuous and differentiable, we reiterate that Sklar's Theorem is completely general, applying to all joint distributions. In the case of discrete distributions, an open question is whether the conventional marginal-and-conditional approach may be more appropriate. In many cases (discrete or otherwise), the analyst relies on the expert's understanding of the system's causal structure, in which case the assessment of marginal and conditional approach would appear to be more suitable. If, in addition, the random variables are discrete with relatively few possible outcomes, it may be much more straightforward to assess conditional distributions. For example, a risk-analysis model of a nuclear reactor might consider the possible failure of a number of subsystems as a prelude to failure of the containment vessel and release of nuclides into the environment. The state of a subsystem may be modeled as a discrete variable (e.g., functional or failed), and the probability of failure of the containment vessel would likely be assessed conditional on the state of the subsystems, the assessments being informed by the expert's understanding of causal relationships in the system. A counterexample, however, might be that the expert could consider the state of several subsystems and, rather than base assessments on causal reasoning, might for convenience assess the marginal probability of failure for each subsystem and the correlations among them.

Correlation assessment is a key element in the use of the copula approach. Much research must be done to understand what assessment methods work best, under what conditions, and what biases may come into play. Although our empirical results are far from conclusive, we hasten to add that errors in probability assessment are well understood and rampant, yet probability assessments are made and used often. Rather than asking whether experts can assess correlations accurately, perhaps we should ask whether they can assess correlations well enough to be useful in the modeling process. The results we report here, although preliminary, suggest an affirmative answer.

Good decision and risk analysis does incorporate thinking about relationships among variables. Conventionally, this is done through careful thinking about conditional relationships, often aided by causal reasoning on the part of the expert. This approach has proven itself as a useful way to cope with complex knowledge elicitation and modeling situations. Use of correlations and copulas provides a useful and practical alternative tool for analysts.[7]

[7] We thank Greg Fischer, George MacKenzie, and Bob Winkler for many discussions on both general and technical issues relating to the assessment of correlations and the use of copulas in decision analysis. Scott Ferson, Jim Smith, and Bob Winkler provided useful comments on various drafts of this paper. This work was supported in part by the Board of Research at Babson College and by the National Science Foundation under Grant SBR 95-96176.

References

Apostolakis, G., S. Kaplan 1981. Pitfalls in Risk Calculations. *Reliability Engrg.* **2** 135–145.

Barlow, R. E., F. Proschan 1975. *Statistical Theory of Reliability and Life Testing: Probability Models*. Holt, Rinehart, and Winston, New York.

Beneš, V., J. Štephán 1997. *Distributions with Given Marginals and Moment Problems*. Kluwer, Dordrecht, Netherlands.

Burmaster, D., P. D. Anderson 1994. Principles of Good Practice for the Use of Monte Carlo Techniques in Human Health and Ecological Risk Assessments. *Risk Anal.* **14** 477–481.

Clemen, R. T. 1996. *Making Hard Decisions: An Introduction to Decision Analysis*, 2nd ed. Duxbury, Belmont, CA.

Cooke, R. 1995. *UNICORN: Methods and Code for Uncertainty Analysis*. AEA Technologies, Warrington, UK.

Dall'Aglio, G., S. Kotz, G. Salinetti 1991. *Advances in Probability Distributions with Given Marginals*. Kluwer, Dordrecht, Netherlands.

DeVuyst, E. A., P. V. Preckel, S. Liu 1999. Discrete Approximations of Joint Probability Distributions. Draft, University of Illinois.

Edgeworth, F. Y. 1898. On the Representation of Statistics by Mathematical Formulae. *J. Royal Statist. Soc.* **61** 670–700.

Frank, M. J. 1979. On the Simultaneous Associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Math.* **19** 194–226.

Frees, E. W., J. Carriere, E. A. Valdez 1996. Annuity Valuation with Dependent Mortality. *J. Risk Insurance* **63** 229–261.

——, E. A. Valdez 1998. Understanding Relationships Using Copulas. *North Amer. Actuarial J.* **2** 1–25.

Genest, C., J. MacKay 1986. The Joy of Copulas: Bivariate Distributions with Uniform Marginals. *Amer. Statist.* **40** 280–283.

Gigerenzer, G. 1991. How to Make Cognitive Illusions Disappear: Beyond Heuristics and Biases. *European Rev. Social Psych.* **2** 83–115.

——, U. Hoffrage, H. Kleinbolting 1991. Probabilistic Mental Models: A Brunswikian Theory of Confidence. *Psychological Rev.* **98** 506–528.

Gokhale, D. V., S. J. Press 1982. Assessment of a Prior Distribution for the Correlation Coefficient in a Bivariate Normal Distribution. *J. Royal Statist. Soc. (A)* **145** 2 237–249.

Howard, R. A., J. Matheson 1983. *The Principles and Applications of Decision Analysis*. SDG, Inc., Menlo Park, CA.

Jouini, M., R. T. Clemen 1996. Copula Models for Aggregating Expert Opinions. *Oper. Res.* **44** 444–457.

Kadane, J. B., J. Dickey, R. L. Winkler, W. Smith, S. Peters 1980. Interactive Elicitation of Opinion for a Normal Linear Model. *J. Amer. Statist. Assoc.* **75** 845–854.

Keefer, D. 1994. Certainty Equivalents for Three-Point Discrete-distribution Approximations. *Management Sci.* **40** 760–773.

——, S. E. Bodily 1983. Three-Point Approximations for Continuous Random Variables. *Management Sci.* **29** 595–609.

Kelly, K. S., R. Krzysztofowicz 1996a. A Bayesian Model and Choice of Expert. Draft, University of Virginia. Charlottesville, VA.

——, —— 1996b. A Bayesian Model of Multiple Experts. Draft, University of Virginia. Charlottesville, VA.

Kraan, B., R. Cooke 1997. The Effect of Correlations in Uncertainty Analysis: Two Cases. R. Cooke, ed. *Technical Committee Uncertainty Modeling: Report on the Benchmark Workshop Uncertainty/Sensitivity Analysis Codes*. European Safety and Reliability Association, Delft, Netherlands.

Kruskal, W. 1958. Ordinal Measures of Association. *J. Amer. Statist. Assoc.* **53** 814–861.

Lacke, C. 1998. *Decision Analytic Modeling of Colorectal Cancer Screening Policies*. Ph.D. Dissertation, North Carolina State University.

Lehmann, E. L. 1966. Some Concepts of Dependence. *Ann. Math. Statist.* **37** 1137–1153.

MacKenzie, G. R. 1994. *Approximately Maximum-Entropy Multivariate Distributions with Specified Marginals and Pairwise Correlations*. Ph.D. Dissertation, University of Oregon.

Mardia, K. V. 1970. A Translation Family of Bivariate Distributions and Fréchet's Bounds. *Sankhya*, Series A, **32** 119–122.

Meeuwissen, A. M. H. 1993. *Dependent Random Variables in Uncertainty Analysis*. Ph.D. Dissertation, Technische Universiteit Delft.

——, R. Cooke 1994. Tree Dependent Random Variables. Report 94-28, Department of Mathematics, Technische Universiteit Delft.

Morgan, M. G., M. Henrion 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge.

Nelsen, R. B. 1986. Properties of a One-Parameter Family of Bivariate Distributions with Specified Marginals. *Comm. Statist.—Theory and Methods* **15** 3277–3285.

—— 1991. Copulas and Association. G. Dall'Aglio, S. Kotz, G.

Salinetti, eds. *Advances in Probability Distributions with Given Marginals*. Kluwer, Dordrecht, Netherlands. 51–74.

—— 1995. Copulas, Characterization, Correlation, and Counterexamples. *Math. Magazine* **68** 193–198.

Reilly, T. 1998. Sensitivity Analysis for Dependent Variables. Draft, Babson College.

Runde, A. S. 1997. *Estimating Distributions, Moments, and Discrete Approximatioins of a Continuous Random Variable Using Hermite Tension Splines*. Ph.D. Dissertation, University of Oregon.

Schweizer, B. 1991. Thirty Years of Copulas. G. Dall'Aglio, S. Kotz, G. Salinetti, eds. *Advances in Probability Distributions with Given Marginals*, Kluwer, Dordrecht, Netherlands. 13–50.

——, E. F. Wolff 1981. On Non-parametric Measures of Dependence for Random Variables. *Ann. Statist.* **9** 879–885.

Sklar, A. 1959. Fonctions de Répartition à *n* Dimensions et Leurs Marges. Publications de l'Institut Statistique de l'Université de 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** 467–474.

Winkler, R. L., W. Smith, R. Kulkarni 1978. Adaptive Forecasting Models based on Predictive Distributions. *Management Sci.* **24** 977–986.

Yi, W., V. M. Bier 1998. An Application of Copulas to Accident Precursor Analysis. *Management Sci.* **44** S257–S270.

Accepted by L. Robin Keller; received March 20, 1997. This paper has been with the authors 6 months for 2 revisions.