Defining copulas

Copulas – a function for creating a joint probability distribution for two or more marginal distributions – have become part of the derivatives pricing mainstream. In this article, the first in a series explaining recent developments in risk management and derivatives pricing, ***John Hull*** *gives a brief explanation of Gaussian, Student-t and multivariate copulas*

John Hull

Copulas have become increasingly significant in risk management and derivatives valuation in recent years. They are an important way of quickly defining a correlation structure between two or more variables. Recently, they have been used extensively in the valuation of collateralised debt obligations and other credit derivatives.

But how do copulas work? Consider two correlated variables V_1 and V_2. The marginal distribution of V_1 (sometimes also referred to as the unconditional distribution) is its distribution assuming we know nothing about V_2; similarly, the marginal distribution of V_2 is its distribution assuming we know nothing about V_1. Suppose we have estimated the marginal distributions of V_1 and V_2. How can we make an assumption about the correlation structure between the two variables to define their joint distribution?

If the marginal distributions of V_1 and V_2 are normal, a convenient and easy-to-work-with assumption is that the joint distribution of the variables is bivariate normal.[1] Similar assumptions are possible for some other marginal distributions, but often there is no natural way of defining a correlation structure between two marginal distributions. This is where copulas come in.

The formal mathematical definition of a copula leaves most people cold. In essence, a copula is nothing more than a way of creating a joint probability distribution for two or more variables while preserving their marginal distributions. The joint probability of the variables of interest is defined implicitly by mapping them to other variables that have a known joint distribution.

Explaining how copulas work can be achieved using a simple example. Suppose the marginal distributions of V_1 and V_2 are the triangular probability density functions shown in figure 1 (triangular distributions make good examples because it is easy to

[1] *Although this is a convenient assumption, it is not the only one that can be made. There are many other ways in which two normally distributed variables can be dependent on each other. For example, we could have* $V_1 = V_2$ *for* $V_2 < k$ *and* $V_1 = -V_2$ *for* $V_2 \geq k$

do calculations for them). Both variables have values between zero and one. The density function for V_1 peaks at 0.2. The density function for V_2 peaks at 0.5. For both density functions, the maximum height is 2.0.

In a Gaussian copula, the variables of interest are mapped to normally distributed variables and new variables are assumed to be multivariate normal. In our example, we map V_1 and V_2 into new variables U_1 and U_2 that have standard normal distributions (a standard normal distribution is a normal distribution with mean zero and standard deviation one).

The mapping is accomplished on a percentile-to-percentile basis. The one-percentile point of the V_1 distribution is mapped to the one-percentile point of the U_1 distribution. The 10-percentile point of the V_1 distribution is mapped to the 10-percentile point of the U_1 distribution, and so on. V_2 is mapped into U_2 in a similar way.

Table A shows how values of V_1 are mapped into values of U_1. Table B similarly shows how the values of V_2 are mapped into values of U_2. Consider the $V_1 = 0.1$ calculation in table A. The cumulative probability that V_1 is less than 0.1 is (by calculating areas of triangles) 5%. The value 0.1 for V_1 therefore gets mapped to the five-percentile point of the standard normal distribution, or –1.64.

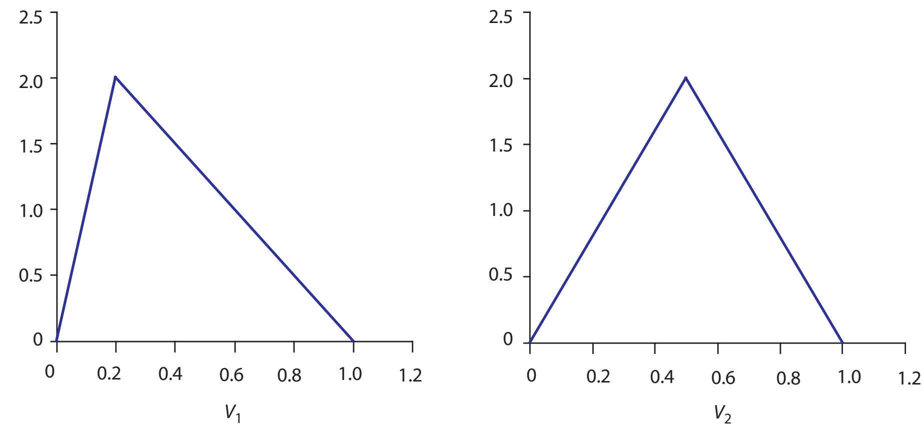
U_1 and U_2 have normal distributions. We assume that they are jointly bivariate normal. This, in turn, implies a joint distribution and a correlation structure between V_1 and V_2. The essence of copula is therefore that, instead of defining a correlation structure between V_1 and V_2 directly, we do so indirectly.

We map V_1 and V_2 into other variables that have 'well-behaved' distributions and for which it is easy to define a correlation structure.

The way in which a copula defines a joint distribution is illustrated in figure 2. Suppose we assume that the correlation between U_1 and U_2 is 0.5. To calculate the joint cumulative probability distribution between V_1 and V_2, we use the bivariate normal distribution.

Suppose we want to know the probability that $V_1 < 0.1$ and $V_2 < 0.1$. From tables A and B, this is the same as the probability that $U_1 < -1.64$ and $U_2 < -2.05$.

1 Probability densities for V_1 and V_2



From the cumulative bivariate normal distribution, this is 0.006 when $\rho = 0.5$.[2] (The probability would be only $0.02 \times 0.05 = 0.001$ if $\rho = 0$). The 0.5 correlation between U_1 and U_2 is referred to as the copula correlation. This is not, in general, the same as the coefficient of correlation between V_1 and V_2.

Other copulas

The Gaussian copula is just one copula that can be used to define a correlation structure between variables. There are many other copulas, leading to many other correlation structures. One that is sometimes used is the Student-*t* copula. Here, the variables of interest are mapped to new variables that are assumed to have a Student-*t* distribution. The mapping is fractile-to-fractile, as for the Gaussian copula.

Figure 3 shows plots of 5,000 random samples from a bivariate normal and the bivariate Student-*t*. The correlation parameter is 0.5 and the number of degrees of freedom for the Student-*t* is four. Define a 'tail value' of a distribution as a value in the left or right 1% tail of the distribution.

There is a tail value for the normal distribution when the variable is greater than 2.33 or less than –2.33.

Similarly, there is a tail value in the *t* distribution when the value of the variable is greater than 3.75 or less than –3.75. Vertical and horizontal lines in the figures indicate when tail values occur. The figures illustrate that it is more common for both variables to have tail values in

[2] *An Excel function for calculating the cumulative bivariate normal distribution is on the author's website:* www.rotman.utoronto.ca/~hull

A. Mapping of V_1 to U_1

V_1 value	Percentile of distribution	U_1 value
0.1	5.00	–1.64
0.2	20.00	–0.84
0.3	38.75	–0.29
0.4	55.00	0.13
0.5	68.75	0.49
0.6	80.00	0.84
0.7	88.75	1.21
0.8	95.00	1.64
0.9	98.75	2.24

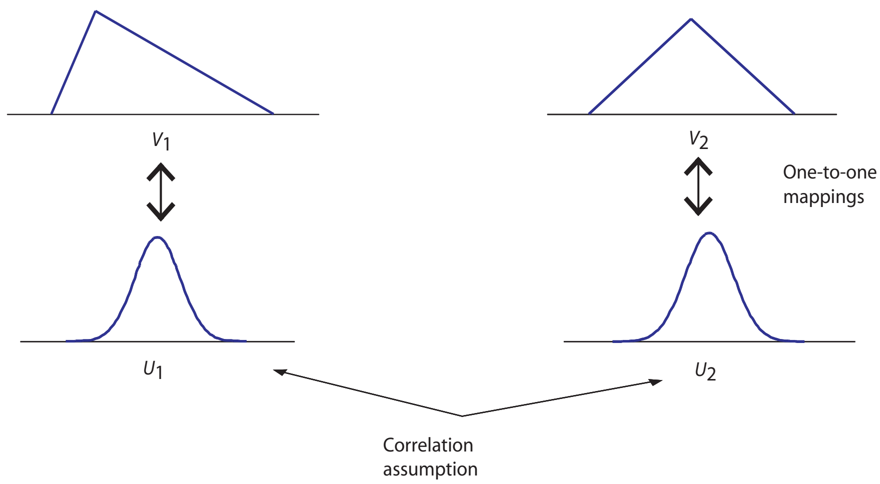
B. Mapping of V_2 to U_2

V_2 value	Percentile of distribution	U_2 value
0.1	2.00	–2.05
0.2	8.00	–1.41
0.3	18.00	–0.92
0.4	32.00	–0.47
0.5	50.00	0.00
0.6	68.00	0.47
0.7	82.00	0.92
0.8	92.00	1.41
0.9	98.00	2.05

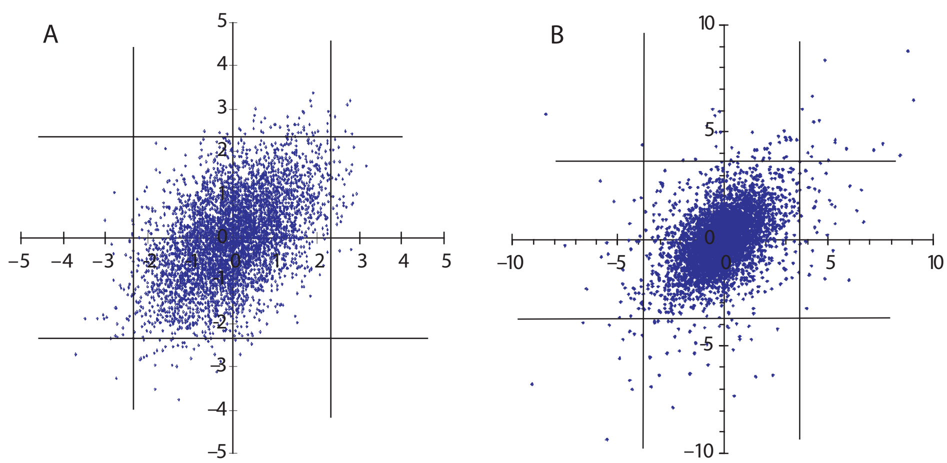
the bivariate *t* distribution than in the bivariate normal distribution.

To put it another way, the tail correlation is higher in a bivariate *t* distribution than in a bivariate normal distribution. Often, the correlation between market variables tends to increase in extreme market conditions. As a result, it can be argued that a Student-*t* copula provides a better description of the behaviour of many market variables than the Gaussian copula.

2 How a copula model defines the joint distribution



3 5,000 random samples from a) a bivariate normal distribution and b) a bivariate Student-*t* distribution



Multivariate copulas

Copulas can be used to define a correlation structure between more than two variables. The simplest example of this is the multivariate Gaussian copula. Suppose that there are N variables, V_1, V_2, ... , V_N and that we know the marginal distribution of each variable. For each i $(1 \le i \le n)$, we transform V_i into U_i, where U_i has a standard normal distribution. (As described above, the transformation is accomplished on a percentile-to-percentile basis). We then assume that the U_i have a multivariate normal distribution.

In multivariate copula models, analysts often assume a factor model for the correlation structure between the Us. When there is only one factor, we can create a Gaussian copula by assuming:

$$U_i = a_i F + \sqrt{1 - a_i^2}\, Z_i$$

where the F and the Z_i have standard normal distributions. The Z_i are uncorrelated with each other and uncorrelated with F. Other factor copula models are obtained by choosing F and the Z_i to have other zero-mean unit-variance distributions. For example, if Z_i is normal and F has a Student-*t* distribution, we obtain a multivariate Student-*t* distribution for the U_i. These distributional choices affect the nature of the dependence between the variables.

Conclusions

An important application of multivariate factor copula models is to model defaults on portfolios when credit derivatives are evaluated. In this case, V_i is the time to default for the ith company in the portfolio. As derivatives where there are two or more underlying variables become more commonplace, it is likely that copulas will play a bigger role in the valuation of derivatives. Also, as risk management models become more sophisticated, it is likely that copulas will be increasingly involved in defining the correlation structure between market variables. ■

This is an edited extract from John Hull's new book, *Risk Management and Financial Institutions*, published by Prentice Hall in 2006, *www.rotman.utoronto.ca/~hull*

References

Cherubini U, E Luciano and W Vecchiato, 2004
Copula methods in finance
Wiley

Demarta S and A McNeil, 2005
The t-copula and related copulas
International Statistical Review 73(1), pages 111–129

Hull J, 2006
Risk management and financial institutions
Upper Saddle River, NJ, Prentice Hall