A Practical Concept of Tail Correlation

B. John Manistre[1], FSA, FCIA, MAAA

[1] The author is Vice President for Risk Research in the Group Risk Dept. of AEGON NV, 1111 North Charles St., Baltimore, Md.

Abstract[2]

This paper shows how the results of copula based capital aggregation models can always be locally approximated by relatively simple formulas. The paper defines the concepts of diversification factor and tail correlation matrix and describes methods for estimating these quantities from simulated data. We show how these ideas can be put into practice as both computational shortcuts and presentation tools. Some examples are then developed, which suggest that, when copula based models are used to aggregate capital, two new phenomena emerge: a) diversification benefits are reduced because of additional tail dependence in the copula; and b) diversification benefits are increased when aggregating risks that have finite variance and the model does not have too much symmetry. Since few of the risks held by a life insurer are so heavy-tailed that they have infinite variance, the paper concludes by arguing that simple, correlation matrix-based, capital aggregation formulas are more defensible than previously thought.

[2] The views and opinions expressed in this paper are those of the author and not AEGON NV.

Introduction

This paper discusses the top-down economic capital aggregation process used by many financial institutions with a wide variety of risks, business units and geographic territories. Assuming we have *n* risks $X_1,...,X_n$, and we have determined stand-alone capital requirements $c_1,...,c_n$ for each risk, then the problem is to determine a reasonable capital requirement $C = C(c_1,...,c_n)$ for the aggregate risk $X = \sum_i X_i$. A simple, and widely used, approach is to choose a correlation matrix ρ_{ij} and assume that we can use the formula:

$$C(c_1,...,c_n) = \sqrt{\sum_{i,j} \rho_{ij} c_i c_j}$$

to get the required result.

Two common criticisms of this approach are:

1. The model is too simple. If we use "ordinary" correlations in the formula, we may not capture the tail behavior of the risks appropriately.

2. The model assumes capital can be moved easily across business unit, legal entity or geographic boundaries.[3]

The main point of this paper is to argue that the first problem is not as severe as it might first appear. The second problem is, strictly speaking, outside the scope of the current paper other than to note that, if we can dispose of Problem 1, then Problem 2 is the real issue.

The main steps of this paper's argument are as follows:

1. The standard aggregation approach makes two, potentially offsetting, theoretical errors.

    a. It ignores the additional correlation that may be present in extreme events as articulated in Problem 1.

    b. It also ignores an additional diversification benefit that can arise when aggregating risks that are not too heavy-tailed in a sense to be made more precise later.

Section 1 of this paper is devoted to developing this argument.

2. As a practical matter, many of the risks faced by a financial institution are not very heavy-tailed. One possible exception is operational risk. The bulk of the evidence to support this statement is cited in Appendix 3. The two theoretical

[3] See, for example, Filipovic, D., and Kupper, M., "Optimal Capital and Risk Transfers for Group Diversification." Preprint dated July 2006.

issues in point (1) above will therefore tend to offset when we build more complex aggregation models. Numerical examples presented in Section 2 of this paper support this point of view.

3. As a practical matter, many of the parameters used in aggregation models are not known with a great deal of precision. As we build more complex models, we must consider more, poorly known, parameters. Unless we can overcome the model/parameter uncertainty issues, it doesn't make sense to make the models too complicated. Furthermore, one of this paper's technical results is that we can always locally approximate a complex model with a simple one as described below.

Apart from the high level argument outlined above, this paper also develops a number of technical tools that anyone doing quantitative work in this area should find useful

1. Under the reasonable assumption that the true capital aggregation formula satisfies the scaling property $C(\lambda c_1,...,\lambda c_n) = \lambda C(c_1,...,c_n)$, we show that there is always a family of local formula approximations, valid in a neighborhood of the point $c^0{}_1,...,c^0{}_n$, of the form:

$$C \approx \sum_i D_i c_i \qquad D_i = \frac{\partial C}{\partial c_i}(c^0{}_1,...,c^0{}_n) \tag{1}$$

$$C \approx \sqrt{\sum_{i,j} D_{ij} c_i c_j} \qquad D_{ij} = \frac{1}{2}\frac{\partial^2 C^2}{\partial c_i \partial c_j}(c^0{}_1,...,c^0{}_n) \tag{2}$$

For a function with the assumed scaling property, the first formula approximation is equivalent to a first order Taylor expansion about any point. In this paper we call the quantities D_i the *diversification factors*.

Under the same scaling assumption, the second formula has second order contact with exact capital function. We will call the matrix D_{ij} the *tail correlation* matrix at the point where it was calculated. The tail correlation matrix is equal to the ordinary linear correlation matrix if the underlying risk model is elliptical[4] in the sense of McNeil, Frey and Embrechts (MFE). We hope the examples presented in this paper will convince the reader that the tail correlation matrix is a useful tool for presenting results.

[4] McNeil, A.J., Frey, R., and Embrechts, P. 2005. *Quantitative Risk Management*. Princeton. Section 3.3 develops the basic theory of elliptical distributions. The multivariate normal and multivariate Student's-t distributions are both examples of elliptical models.

This tail correlation formula looks just like the one used in the standard correlation matrix approach, but there are some important differences.

- In general, the formula (2) above is local in the sense that it is only valid in some neighborhood of the point where it was determined. The examples presented later in the paper suggest this is not an obstacle to practical application.

- The diagonal elements of the tail correlation matrix need not be equal to one, and the off diagonal elements are not constrained to lie between +1 and -1. We interpret this phenomenon later to mean that, in the tail of a loss distribution, risks can be more or less correlated with themselves. Depending on one's point of view, this is either an intriguing idea or a presentation nightmare. One of this paper's objectives is to help the reader get comfortable with this broader concept of a correlation matrix.

2. The paper presents tools to estimate the diversification factors and tail correlation matrix from real or simulated data. Two methods of estimation are described, which we call the direct and indirect methods, respectively. Using these tools can be very easy in some circumstances, but, when the risks are heavy-tailed, a large number of samples can be required to get accurate estimates. In general, estimating the tail correlation matrix requires more data than estimating the diversification factors for the same level of precision.

3. The paper also develops a variation on traditional extreme value theory. For risks with finite variance, we define the *tail shape function* $\xi(u)$ by:

$$1-2\xi(u)=\frac{[CTE(u)-VaR(u)]^2}{CTV(u)},\quad 0<u<1$$

where $CTE(u)$ is the conditional tail expectation at probability level u; $VaR(u)$ is the value at risk; and $CTV(u)$ is the conditional tail variance. If we know the tail shape function for all u, then we can determine the risk up to location and scale. For a Pareto distribution, this quantity is a constant. The example in Section 2 shows that if different marginal models are parameterized to have the same location, scale and tail shape behavior for u in a neighborhood below, say .995, then they will all produce similar results when estimating capital measures such as *VaR*(.995) or *CTE*(.99) for the aggregate risk. The key point here is that it is the behavior of the tail shape below a certain threshold that drives the aggregation results. The theory underlying this result is detailed in Appendix 3.

The main implication of this result is that we do not need to know every detail about the marginal probability distribution of the component risks. Knowledge of location, scale and the tail shape function in a neighborhood below a point near $u=.995$ appear to be the most important properties.

The remainder of this paper is structured as follows. Following this introduction is a section devoted to explaining the tool development in more detail. The section starts by looking at sufficient conditions for tail correlation to equal ordinary correlation and then goes on to look at an example of a capital aggregation problem, which is simple enough to be studied in closed form but is not elliptical. This provides a simple laboratory for developing tail correlation intuition. The section then introduces the tools needed to estimate tail correlation from simulated data and provides some examples.

The second section of the paper develops the model building scenario described earlier. This is the core of the paper. A reader willing to accept most of the results in the tool development section can understand this section after skimming the examples in Section 1. We walk through a model building scenario starting with a simple, elliptical, model where tail correlation and ordinary correlation coincide, and then, in several steps, add complexity in the form of more realistic marginal distributions and a more conservative dependency structure. The example shows that making the dependency structure more conservative does make the end result more conservative, but this conservatism can be offset by an additional diversification benefit that arises from the assumption that the individual component risks have a variety of different tail behaviors. We actually have to work very hard to come up with an answer that is more conservative than the starting point. The section ends by asking how well we know the parameters being used and whether the increase in model complexity is genuinely justified.

The third section of the paper develops the conclusions in more detail. The author’s first conclusion is that, unless we think we know the models and parameters extremely well, it is hard to justify introducing a very complex model. The simple correlation model is capable of doing a better job than it may have been given credit for. The second conclusion is that, if we do know how to improve the model, or simply want to add some conservatism, any additional complexity can be approximated by a tail correlation matrix that looks like an ordinary correlation matrix, except for the fact that it may not have ones on the diagonal.

The fourth and final section contains a number of technical appendices related to tool development. These appendices are intended for an audience of specialists who might wish to reproduce some of the results reported in this paper.

1. Tail Correlation as a Presentation Tool

1.1 Homogeneity and the Law of Large Numbers

A function $C = C(c_1,...,c_n)$ is said to be homogeneous of degree 1 if it satisfies the scaling law $C(\lambda c_1,...,\lambda c_n) = \lambda C(c_1,...,c_n)$ for all $\lambda > 0$. The standard formula $C = \sqrt{\sum_{i,j} \rho_{ij} c_i c_j}$ is an example of such a function. Since this paper will assume that all capital aggregation models obey this rule, we briefly indicate why this is reasonable.

Suppose we have sold one-year term insurance contracts to N identical and independent lives. Assume the mortality rate Q for each member of this group of lives is itself a random variable with mean $E[Q] = q$. The total death claims to be experienced over the year are just the sum of N independent Bernoulli risks. The expected number of claims is just:

$$E[D] = Nq,$$

while the variance of the claim count distribution can be calculated from:

$$\begin{aligned} VAR[D] &= E[VAR[D \mid Q]] + VAR[E[D \mid Q]], \\ &= E[NQ(1-Q)] + VAR[NQ], \\ &= Nq(1-q) + N(N-1)VAR(Q). \end{aligned}$$

If we take the number of contracts to be a measure of "exposure" and we take standard deviation to be a reasonable measure of risk, then we can write:

$$\begin{aligned} Stdev(D) &= N\sqrt{\frac{q(1-q)}{N} + (1-1/N)VAR(Q)} \\ &\approx N\sqrt{VAR(Q)}, \quad as\ N \to \infty. \end{aligned}$$

We see that while the pure fluctuation risk will diversify away, any risk associated with the claim rates themselves simply grows with the exposure. The assumption of homogeneity is therefore roughly equivalent to the assumption that the law of large numbers can be applied, and there is no more risk that can be diversified away simply by scaling up the block of business. This is a reasonable assumption for a large insurer with sound risk management practices, such as the appropriate use of reinsurance, to limit the effect of large individual contracts.

Once the assumption of homogeneity has been made, it has a number of useful mechanical consequences. Starting with the scaling assumption $C(\lambda c_1,...,\lambda c_n) = \lambda C(c_1,...,c_n)$, we can differentiate with respect to the parameter λ and then set $\lambda = 1$. The result is:

$$\frac{d}{d\lambda} C(\lambda c_1,...,\lambda c_n) = C(c_1,...,c_n) \Rightarrow \sum_i \frac{\partial C}{\partial c_i} c_i = C,$$

A result which MFE call the Euler Decomposition or allocation of the aggregated capital back to its component risks.

This result can be used to derive a first order formula approximation. Suppose $\tilde{c}_i$ is a point in a neighborhood of c_i, then Taylor's expansion gives:

$$\begin{aligned}
C(\tilde{c}_i) &= C(c_i) + \sum_i \frac{\partial C}{\partial c_i}(\tilde{c}_i - c_i) + \ldots \\
&= \sum_i \frac{\partial C}{\partial c_i}\tilde{c}_i + [C(c_i) - \sum_i \frac{\partial C}{\partial c_i} c_i] + \ldots \\
&= \sum_i D_i \tilde{c}_i + [0] + \ldots
\end{aligned}$$

The homogeneity assumption means the term in square brackets drops out. Thus the homogeneous formula approximation $C \approx \sum_i D_i c_i$ is equivalent to a first order Taylor expansion.

This result could be used to justify factor based capital formulas, like the first generation of capital models, introduced in the early 1990s.

If we want a homogeneous formula approximation accurate to second order, a reasonable place to start is a formula of the form $C \approx \sqrt{\sum_{i,j} D_{ij} c_i c_j}$. If this formula is to have second order contact, then the relation $C^2 \approx \sum_{i,j} D_{ij} c_i c_j$ must also hold to second order. From this we deduce that, if the formula works at all, then we must have $D_{ij} = \frac{1}{2}\frac{\partial^2 C^2}{\partial c_i \partial c_j}$. There are some additional requirements for this formula to be valid; for example, the function C must be positive (not a problem here), suitably smooth, and the matrix $D_{ij} = \frac{1}{2}\frac{\partial^2 C^2}{\partial c_i \partial c_j}$ should be positive semi-definite.

In Appendix 1 we show that, if the risk measure being used is coherent in the sense of Artzner[5] et al., then the tail correlation matrix is guaranteed to be positive semi-definite. Assuming this to be the case, one can then do some calculus to verify that the proposed formula approximation really does have second order contact at the point c_i.

We conclude that, at least for coherent risk measures, we can always find simple formula approximations to an exact capital aggregation model.

[5] Artzner, P., "Application of Coherent Risk Measures to Capital Requirements in Insurance," *North American Actuarial Journal* (April 1999).

1.2 Elliptical Models

Elliptical models are discussed here briefly because they are basically the answer to the question: When is the standard aggregation formula $C = \sqrt{\sum_{i,j} \rho_{ij} c_i c_j}$ actually correct? As detailed in MFE,[6] a random vector X is elliptical if it can be written as an affine function of a spherical vector S.

More formally, X is elliptical if we can write $X = m + A\,R\,S$ where $m = E[X]$ is a constant vector, A is a constant matrix, S is a vector uniformly distributed on an (n-1) sphere, and R is a positive scalar independent of S.

Two important examples of elliptical models are the multivariate normal distribution and the multivariate Student-t model. For the multivariate normal distribution, R is the square root of a χ_n^2 variate while the Student-t model results if $R = \sqrt{n / \chi_n^2}$.

An important property of elliptical models is that if the vector $X = \sum_i X_i$ is elliptical and ζ is a well defined risk measure, such as value at risk (*VaR*) or conditional tail expectation (*CTE*), then we can always write:

$$\zeta(X) = E[X] + \sqrt{\sum_{i,j} \rho_{ij} (\zeta(X_i) - E[X_i])(\zeta(X_j) - E[X_j])}$$

where $\rho_{ij} = \dfrac{\sum_k A_{ik} A_{jk}}{\sqrt{\sum_k (A_{ik})^2 \sum_l (A_{jl})^2}}$. If the vector X has finite variance then the matrix ρ_{ij} is also the usual linear correlation matrix.

In an economic capital context it is usual to assume that all variables have a mean of zero so that the risk measure aggregation formula becomes

$$\zeta(X) = \sqrt{\sum_{i,j} \rho_{ij} \zeta(X_i) \zeta(X_j)} \, .$$

This result follows from a property of spherical distributions proved in MFE.[7] They show that for any constant vector a the linear combination $Ra \cdot S$ is equal, in distribution, to $|a| RS_1$. This means that there is a constant c, depending on the risk measure and R, such that for each i:

$$\zeta(X_i) = m_i + c\sqrt{\sum_k A_{ik}^2} \, .$$

---

[6] McNeil, Frey & Embrechts, as in footnote 2.
[7] McNeil, Frey & Embrechts, Theorem 3.19.

For the aggregate risk we then have:

$$\begin{aligned}\zeta(X) &= E[X] + c\sqrt{\sum_k (\sum_i A_{ik})^2}, \\ &= E[X] + \sqrt{\sum_k c^2 \sum_i A_{ik} \sum_j A_{jk}}, \\ &= E[X] + \sqrt{\sum_{ij} c^2 \sum_k A_{ik} A_{jk}}, \\ &= E[X] + \sqrt{\sum_{ij} \rho_{ij} (\zeta(X_i) - m_i)(\zeta(X_j) - m_j)}.\end{aligned}$$

as required.

Elliptical models therefore have the property that the tail correlation matrix D_{ij} is independent of both the specific choice of risk measure and mix of risks used to calculate it. This is a result of the assumed symmetry in the model, so one would not expect these properties to be very representative.

1.3 A Non-Elliptical Closed Form Example

Let $\xi > 0$ be a real number and consider the aggregation formula $C = [\sum_i c_i^{1/\xi}]^{\xi}$. We will examine this formula, and some variations on it, in some detail because it represents a model which is not elliptical but is still simple enough that we can find a closed form expression for the capital aggregation process.

This formula applies in at least two moderately realistic situations.

1. If $\xi \geq 1/2$, this is the technically correct formula for aggregating capital when the risks are independent, identical stable distributions with index of stability $\alpha = 1/\xi$. The formula also holds if we use a mixture of such risks.

   A random variable Y has a stable distribution if a sum of n independent copies belongs to the same location/scale family, i.e., $Y_1 + ... + Y_n \stackrel{d}{=} n^{1/\alpha} Y$. The normal distribution is a member of the stable family corresponding to $\alpha = 2$. See Nolan[8] for more background on stable distributions.

2. For any $\xi > 0$, extreme value theory[9] can be used to show this is an approximate formula for aggregating capital when

[8] Nolan, John P., "Stable Distributions: Models for Heavy-Tailed Data," Birkhauser (2007). The first chapter of this book can be downloaded, for free, from John Nolan's Web page.

[9] This is a standard result in extreme value theory. See, for example, Bocker, B., Kluppelberg C., "Operational VaR: A Closed Form Solution," *Risk* Magazine (December 2005).

- the risks are independent compound variables with identical severity distributions that have regularly varying tails with tail index ξ and

- the risk measure being used is sensitive only to the far tail of the loss distribution. VaR or CTE with a high percentile choice would be appropriate examples.

If a risk has tail index ξ, this means that there are constants A, B so that a Pareto approximation:

$$VaR(u) \approx A + B(1-u)^{-\xi}$$

is valid in the limit as $u \to 1$. This is the core insight of extreme value theory (see MFE Ch. 7).

If $\xi < 0$, the risk is bounded above and is referred to as light-tailed.

If $\xi < 1/2$, the variance is finite, and, as we show in Appendix 3, the tail index ξ can be defined by the limiting value:

$$1 - 2\xi = \lim_{u \to 1} \frac{[CTE(u) - VaR(u)]^2}{CTV(u)}.$$

An important example in this category is the Student-t distribution with n degrees of freedom. This risk has a tail index of $\xi = 1/n$.

If $\xi > 1/2$ then the risk is considered to be very heavy-tailed. Examples of very heavy-tailed risks are the stable distributions for $\alpha > 2$. In this situation, the relationship between index of stability and tail index is just $\alpha = 1/\xi$.

The case $\alpha = 2$ requires special treatment. As noted earlier, the stable distributions with $\alpha = 2$ are Gaussian, and they are not heavy-tailed. In fact, the tail index of a Gaussian random variable is just $\xi = 0$ as is the tail index of a lognormal random variable. The mathematical reason for this apparent discontinuity in behavior is beyond the scope of this paper; however, the discontinuity is important to the argument that follows.

With this background, we can now ask how the new formula $C = [\sum_i c_i^{1/\xi}]^{\xi}$ differs from the standard. To address this issue, we start with a two risk model such that $c_1 = c_2$. Table 1.1 below shows the results of applying the ideas developed in the previous section

TABLE 1.1 $c_1 = c_2$

	$\xi =$ 0.35		$\xi =$ 0.50		$\xi =$ 0.65	
D	64%		71%		78%	
D_1, D_2	64%	64%	71%	71%	78%	78%
D_{ij}	116%	-35%	100%	0%	95%	28%
	-35%	116%	0%	100%	28%	95%

The first row in the table shows the aggregate diversification benefit $D = C / \sum_i c_i$, while subsequent rows show the diversification factors and tail correlation matrix defined earlier. In this example, the two risks enter the problem symmetrically so $D = D_1 = D_2$. We see that the aggregate risk is an increasing function of the tail index parameter ξ. This is not a surprise.

When $\xi = 1/2$, the model reduces to the standard elliptical formula, and the tail correlation matrix is just the two dimensional identity matrix. However, when $\xi < 1/2$, the diagonal elements are greater than 1, and the off diagonal elements are negative even though the risks are independent. These relationships are reversed when $\xi > 1/2$.

This example suggests that, in the presence of elliptical symmetry, risks aggregate as if they had a tail index of $\xi = 1/2$ regardless of their actual tail index.

Before trying to draw too many conclusions, it is useful to look at a second example, where the risks are not of the same size. Table 1.2 assumes the first risk is twice the size of the second risk.

TABLE 1.2 $c_1 = 2\, c_2$

	$\xi =$ 0.35		$\xi =$ 0.50		$\xi =$ 0.65	
D	70%		75%		81%	
D_1, D_2	92%	25%	89%	45%	90%	62%
D_{ij}	106%	-20%	100%	0%	96%	26%
	-20%	93%	0%	100%	26%	99%

The first order diversification factors now reflect a fundamental risk principle, the marginal cost of adding exposure to a risk increases with the amount of exposure that you already have. This can be verified by doing some calculus to get:

$$D_i = \frac{\partial C}{\partial c_i} = \left(\frac{c_i}{C}\right)^{1/\xi - 1}.$$

One interesting observation about Example 1.2 is that the tail correlation element D_{22} is now less than 1. To get a better understanding of what is going on, we examine the formula for D_{ij}, which shows that the tail correlation entries are fairly simple functions of the first order diversification factors. By direct calculation we find:

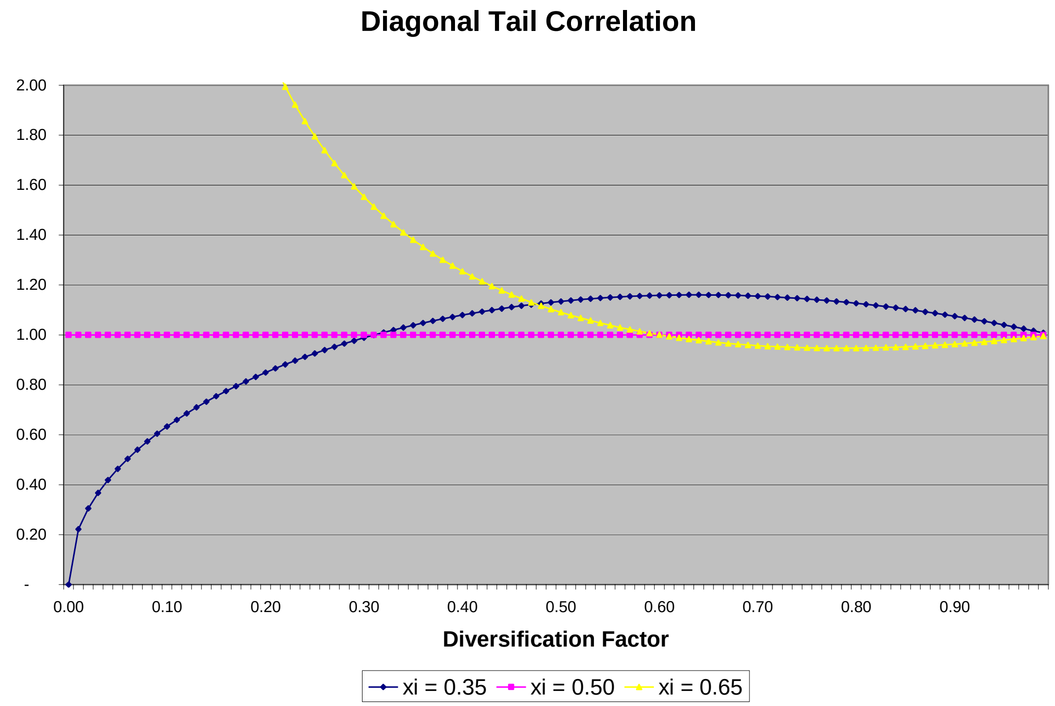
$$D_{ij} = \frac{1-\xi}{\xi} D_j^{\frac{1-2\xi}{1-\xi}} \delta_{ij} - \frac{1-2\xi}{\xi} D_i D_j, \quad \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

From this result, we see the sign of the off diagonal terms is determined only by the tail shape parameter. If $\xi > 1/2$, then the tail correlation is always positive. When $\xi < 1/2$, then the reverse is true.

If we set $i = j$ in the formula above, we find for a diagonal element:

$$D_{ii} = \frac{1-\xi}{\xi} D_i^{\frac{1-2\xi}{1-\xi}} - \frac{1-2\xi}{\xi} D_i^2.$$

The following chart plots the behavior of D_{ii}, as a function of D_i, for three values of the shape parameter ξ.



As the chart shows, the diagonal elements have the potential to take on a wide range of possible values. But extremely large deviations from 1 only occur when the first order diversification factor is small.

It is possible to construct a closed form example that combines the issues in the model above with traditional correlation concepts. The example is fairly easy to understand in the context of stable distributions. Let $Y_A,\ A = 1,\ldots,m$ be m independent, identical, symmetric stable random variables with index of stability $\alpha = 1/\xi,\ \alpha \geq 2$. Now assume there is a positive random variable W, independent of the Y_A and a matrix B_{iA} so that the n risks in our aggregation model are equal, in distribution, to $X_i = W\sum_A B_{iA}Y_A$. By the properties of stability, we must have $X_i \overset{d}{=} b_i WY$ where $b_i = [\sum_A |B_{iA}|^{1/\xi}]^{\xi}$. There will then be a constant $k > 0$, depending on the risk measure, such that $c_i = kb_i$. For the aggregate capital we then have:

$$\begin{aligned} C &= k[\sum_A |\sum_i B_{iA}|^{1/\xi}]^{\xi} \\ &= [\sum_A |\sum_i kB_{iA}|^{1/\xi}]^{\xi} \\ &= [\sum_A |\sum_i \frac{c_i}{b_i} B_{iA}|^{1/\xi}]^{\xi} \end{aligned}$$

While more complex than the earlier formula, this is still simple enough that we can compute diversification factors and tail correlations in closed form. Table 1.3 below revisits the $c_1 = c_2$ example but with the matrix B above chosen so that the two risks have an ordinary correlation of 25% when $\xi = 1/2$.

TABLE 1.3 $c_1 = c_2$

	$\xi =$ 0.35		$\xi =$ 0.50		$\xi =$ 0.65	
D	73%		79%		85%	
D_1, D_2	73%	73%	79%	79%	85%	85%
D_{ij}	109%	-3%	100%	25%	97%	49%
	-3%	109%	25%	100%	49%	97%

The result is, very roughly, to add 25% to the off diagonal terms while moving the self correlations closer to one.

There is one sense in which the examples treated so far are not very representative. If we increase the number of risks involved from 2 to 4, then the individual diversification factors will drop because each risk is, relatively, less important. This has the effect of bringing the off diagonal terms closer to 0 and bringing the diagonal terms closer to one. Table 1.4 illustrates this point for $\xi = .35$.

TABLE 1.4

$\xi =$ 0.35 $c_1 = c_2 = c_3 = c_4$

D	41%			
D_i	41%	41%	41%	41%
D_{ij}	108%	-14%	-14%	-14%
	-14%	108%	-14%	-14%
	-14%	-14%	108%	-14%
	-14%	-14%	-14%	108%

The examples of this section are still too special to allow any firm conclusions to be drawn, but they do suggest some ideas that can be tested with the tools to be introduced in the next section.

A reasonable hypothesis is:

The standard aggregation formula may be conservative when aggregating risks whose tail indices are less than ½.

A counterargument to this hypothesis is that all of the examples in Section 1.3 have a very simple copula structure. A copula assumption that builds in more tail dependence[10] might well overcome the effect identified above, assuming the effect is real to begin with.

In order to address this issue, we need to build a more sophisticated tool that can deal numerically with very general models.

1.4 Some Numerical Examples: Aggregating Pareto Risks with the Gaussian and t-Copulas

In order to test the idea raised at the end of Section 1.3, we need to work with models for which there are no simple closed form expressions for capital aggregation. We start with the problem of aggregating four independent, identical Pareto risks with tail index $\xi = .330$. The standard correlation approach to this problem would produce an aggregate diversification benefit of $D = 50\%$.

A simulation model was developed to generate 100,000 samples of four independent Pareto variates. Using a risk measure of *CTE*(.99) and the direct method, as described in Appendix 1, we estimated the diversification measures described in Section 1.3. The numerical experiment was then repeated 10 times and the results averaged to reduce sampling error. The results are in Table 1.5 below:

TABLE 1.5 $\xi =$ 0.33

D	44.5%				+-	0.1%			
D_i	44%	44%	45%	45%	+-	1%	1%	1%	1%
D_{ij}	100%	-6%	-7%	-8%		1%	1%	1%	1%
	-6%	96%	-5%	-6%	+-	1%	2%	1%	1%
	-7%	-5%	100%	-7%		1%	1%	2%	1%
	-8%	-6%	-7%	101%		1%	1%	1%	3%

The results are similar, but not identical, to Table 1.4 in the previous section. The left-hand side of the table reports the mean of the 10 experiments, while the standard error is reported on the right-hand side. Since all of the risks are entering the problem in a symmetric way, we would expect all of the off diagonal elements in the tail correlation matrix to be equal. They are equal to within the estimated sampling error. The same comment applies to the diagonal elements.

[10] See MFE Chapter 5 for a discussion of copulas and tail dependence.

We now briefly indicate what has to be done in order to compute the results shown in Table 1.5. When the risk measure being using is *CTE*(*u*), it can be shown that the first derivatives[11] of the capital aggregation function $C = C(c_1,...,c_n)$ are given by:

$$D_i = \frac{\partial C}{\partial c_i} = \frac{E[X_i \mid X \geq VaR_X(u)]}{c_i}$$

while the second derivatives must be estimated from the formula:[12]

$$\frac{\partial^2 C}{\partial c_i \partial c_j} = \frac{Cov[(X_i, X_j) \mid X = VaR_X(u)]}{c_i c_j (1-u)} f_X(VaR_X(u)).$$

Here $f_X(VaR_X(u))$ is the probability density of the aggregate risk $X = \sum_i X_i$ when it is equal to its *VaR*(*u*) level.

The tail correlation matrix is then estimated from the formula:

$$D_{ij} = \frac{1}{2}\frac{\partial^2 C^2}{\partial c_i \partial c_j} = D_i D_j + C\frac{\partial^2 C}{\partial c_i \partial c_j}.$$

There are a number of technical challenges that must be overcome before these results can be put into practice. See Appendices 1 and 4 for more details.

When working with risk measures, such as *CTE*, many of the obvious statistical estimators are known to exhibit small sample bias.[13] We believe the sample size of 100,000 is large enough, in this case, to avoid that issue.

We now ask what happens if we remove the assumption that the four risks are independent. We continue to assume the four risks are identical, but we will now assume the dependency structure can be modeled by a Student-t copula with 10 degrees of freedom.[14] In addition to the degrees of freedom parameter, the t-copula requires us to specify the rank correlation matrix of the risks. We continue to assume the rank correlation is simply the identity matrix. The results are in Table 1.6, which reports the diversification metrics and the standard linear correlation.

[11] For a rigorous derivation of this result, see Tasche, D., "Risk Contributions and Performance Measurement" Preprint (2000).
[12] This appears to be an original result. See Appendix 1 for a high level proof and generalizations to other risk measures such as *VaR*.
[13] See, for example, Manistre, B.J. & Hancock, G.H., "Variance of the CTE Estimator," *North American Actuarial Journal* (May 2005).
[14] See MFE Chapter 5 for a background discussion of copulas.

TABLE 1.6 $\xi = 0.330$ $\nu = 10$

D	49.2%				+-	0.1%			
D_i	49%	48%	50%	49%	+-	1%	1%	1%	1%
D_{ij}	105%	-5%	0%	-3%	+-	3%	1%	1%	2%
	-5%	100%	0%	-1%		1%	2%	1%	2%
	0%	0%	96%	2%		1%	1%	3%	1%
	-3%	-1%	2%	99%		2%	2%	1%	3%
ρ_{ij}	100%	4%	3%	4%	+-	0%	0%	0%	1%
	4%	100%	4%	4%		0%	0%	0%	0%
	3%	4%	100%	4%		0%	0%	0%	0%
	4%	4%	4%	100%		1%	0%	0%	0%

The use of the t-copula has clearly made the results more conservative. We see that the diversification benefit that arose from assuming the tail index was less than one-half has been almost completely offset by the additional tail dependence assumed in the copula. The additional tail dependence has also made the tail correlation metric harder to estimate. Table 1.6 uses the same number of samples as Table 1.5, i.e., 10 runs of 100,000 samples.

We close this section with one final example. The tail index assumption of $\xi = .330$ is still fairly heavy for most of the risks in a large company's portfolio. We can make the example a bit more realistic by assuming that half of the risks in the model are lighter, with a tail index of $\xi = .088$.

TABLE 1.7 $\xi_1 = \xi_2 = 0.330$ $\nu = 10$
$\xi_3 = \xi_4 = 0.088$

D	46.8%				+-	0.1%			
D_i	48%	48%	45%	46%	+-	0.00	0.00	0.00	0.00
D_{ij}	99%	4%	-5%	-6%	+-	1%	1%	1%	1%
	4%	98%	-5%	-7%		1%	1%	1%	1%
	-5%	-5%	111%	-15%		1%	1%	2%	1%
	-6%	-7%	-15%	114%		1%	1%	1%	2%
ρ_{ij}	100%	4%	3%	3%	+-	0%	0%	0%	0%
	4%	100%	3%	3%		0%	0%	0%	0%
	3%	3%	100%	2%		0%	0%	0%	0%
	3%	3%	2%	100%		0%	0%	0%	0%

Not surprisingly, the diversification benefit has increased by assuming some of the risks have lighter tails.

2. An Aggregation Modeling Scenario

With the concepts and tools from Section 1 in hand, we now walk through a small, but moderately realistic, modeling scenario. We assume a model company that has determined its stand-alone capital requirements for four risk types. These are:

1. Investment Risk (IR): The company has determined that it needs four units of capital at the *CTE*(.99) level for credit risk.

2. Mismatch Risk (MR): A combination of sensitivity tests and duration calculations has led to the conclusion that 2.5 units of capital are required at the *CTE*(.99) level.

3. Underwriting Risk (UW): Only 2.0 units of capital are required, at the *CTE*(.99) level, based on sensitivity testing.

4. Operational Risk (OR): A high level model suggests 1.5 units of capital are required at the *CTE*(.99) level.

The sum of the stand-alone capital requirements is 10. For the purposes of this section, we will assume these numbers are known with a fair degree of precision.

As a first step in the aggregation process, the company selects a small group of seasoned risk professionals and assigns them the task of coming up with a reasonable correlation matrix. After researching the available data and the risk literature, the group came up with the recommendation below.

	IR	MR	UW	OR
IR	100%	40%	20%	20%
MR	40%	100%	0%	20%
UW	20%	0%	100%	0%
OR	20%	20%	0%	100%

The group's recommendation also came with some hefty caveats. The group believes the precision of its estimated correlation coefficients is no better than +-10%.

Undaunted by this caveat, the company goes ahead and computes the diversification factors and aggregate capital based on this correlation matrix. The result is Table 2.1.

TABLE 2.1 Step 1

c_i	4	2.5	2	1.5
D	66.0%			
D_i	86%	67%	42%	42%
D_{ij}	100%	40%	20%	20%
	40%	100%	0%	20%
	20%	0%	100%	0%
	20%	20%	0%	100%

These results are of interest in themselves. They tell us that credit risk is the most expensive risk, on the margin, because there is more of it and it is positively correlated to all of the other risks. Underwriting risk and operational risk have similar diversification factors even though they are not identical in size. Operational risk is more expensive because it is more highly correlated to other risks than is underwriting risk. All of these conclusions assume the starting correlation matrix is materially correct.

The second step in the company's modeling process was to build an elliptical simulation model that approximates the results of Step 1. The company started by assuming the component risks had a multivariate Student-t distribution with 100 degrees of freedom and a rank correlation matrix equal to the risk group's recommendation. The scale of each component risk was calibrated to reproduce the stated *CTE*(.99) values. The results are in Table 2.2.

TABLE 2.2 Step 2

Multivariate Student t $\nu = 100$

c_i	4	2.5	2	1.5					
D	66.8%				+-	0.1%			
D_i	87%	67%	43%	44%	+-	1%	1%	0%	1%
D_{ij}	101%	41%	21%	22%		1%	1%	1%	1%
	41%	99%	0%	21%	+-	1%	1%	1%	1%
	21%	0%	102%	0%		1%	1%	1%	1%
	22%	21%	0%	102%		1%	1%	1%	1%
ρ_{ij}	100%	42%	21%	21%		0%	0%	0%	0%
	42%	100%	0%	21%	+-	0%	0%	0%	0%
	21%	0%	100%	0%		0%	0%	0%	0%
	21%	21%	0%	100%		0%	0%	0%	0%

We are now using a simulation model. In this case, a run of 25,000 samples was repeated 10 times to build Table 2.2. Since the model is elliptical, we expect tail correlation and linear correlation to be the same. They are to within sampling error.

The small difference between Tables 2.1 and 2.2 is due to the fact that Table 2.2 assumes the rank correlation is given by the risk group's recommendation. This has had a very small impact on the correlation matrix, but the aggregate diversification factor has gone up by almost one percentage point.

In Step 3 of the process, it is argued that the marginal distributions, Student-t with 100 degrees of freedom, do not have heavy enough tails. Our group of seasoned risk professionals is now asked to recommend a better assumption for each component risk. The group is aware of the research, reported in Appendix 3 of this paper, that the crucial quantity that they need to get right is the tail shape of each risk at or near the .995 probability level. They assign each risk to one of the four tail risk categories in the table below.

After much debate, they conclude that IR and MR are medium-tailed; UR is light-tailed; and OR is heavy-tailed.

Class	**Student- t deg. of freedom**	**Tail Shape at 99.5% (Pareto)**	**Log Normal Sigma**
Light Tailed	25	(0.03)	.19
Medium Tailed	10	0.06	.47
Heavy Tailed	3	0.330	1.52
Very Heavy Tailed	1.5	0.67	N/A

A new simulation model is now built that uses the new, risk specific, degree of freedom parameters. The copula is still based on 100 degrees of freedom, but we have had to increase the sample size from 25,000 to 100,000 because we are using risks with a heavier tail index.

Table 2.3 has the results of running a sample size of 100,000 and then repeating the process 10 times.

TABLE 2.3 Step 3

c_i	4	2.5	2	1.5					
ν_i	10	10	25	3		*Copula*	$\nu =$ 100		
	Stu-t	Stu-t	Stu-t	Stu-t					
D	63.4%				+-	0.1%			
D_i	88%	63%	40%	31%	+-	0.4%	0.3%	0.4%	0.4%
D_{ij}	104%	35%	19%	9%	+-	1%	0%	1%	1%
	35%	98%	-1%	12%		0%	1%	0%	0%
	19%	-1%	92%	-4%		1%	0%	1%	1%
	9%	12%	-4%	91%		1%	0%	1%	3%
ρ_{ij}	100%	37%	21%	19%	+-	0%	4%	0%	0%
	37%	100%	0%	17%		4%	0%	0%	2%
	21%	0%	100%	0%		0%	0%	0%	0%
	19%	17%	0%	100%		0%	2%	0%	0%

This result may be surprising because the aggregate result has gone down even though all of the marginal risk assumptions have become more heavy-tailed. Part of the explanation lies in the fact that, by choosing tail behaviors different from the copula, we no longer have elliptical symmetry in the model. The model in Step 3 above is more like the examples in Section 1.2 with $\xi < 1/2$ than the elliptical models in Section 1.1, which behave like models with $\xi = 1/2$.

A skeptic now demands proof that the results of Step 3 are not very sensitive to all the details of the marginal distribution assumption. Our group of seasoned risk professionals now concludes that the Student-t assumption is reasonable for the MR and UW risks but recommends the use of a lognormal assumption for IR and the Pareto distribution for OR.

We briefly detail what this means. When using a Student-t model, we simulate individual risks using a formula for the form $VaR(u) = BT_n^{-1}(u)$, where the scale constant B is chosen to calibrate the model to the desired level of stand-alone capital c_i.

If a Pareto model is assumed, the simulation formula becomes $VaR(u) = A + B(1-u)^{-\xi}$. The three parameters A, B, ξ are now chosen to satisfy three constraints.

- The mean of the distribution is zero.
- The model calibrates to the desired stand-alone capital $CTE(.99)$.
- The model has the same tail shape at the .995 level as the Student-t distribution it is replacing. In practice this means using the Pareto parameter in the preceding table.

For a lognormal model we go through the same parameter fitting process with the simulation formula $VaR(u) = A + B\exp[\sigma\,\Phi^{-1}(u)]$.

The results are in Table 2.4.

TABLE 2.4 Step 4

c_i	4	2.5	2	1.5						
ν_i	10	10	25	3		*Copula*	$\nu=$ 100			
	Lg Nrml	Stu-t	Stu-t	Pareto						
D	61.6%				+-	0.1%				
D_i	87%	61%	38%	27%	+-	0.5%	0.3%	0.4%	0.5%	
D_{ij}	105%	31%	17%	4%	+-	1%	0%	0%	1%	
	31%	91%	2%	9%		0%	0%	0%	0%	
	17%	2%	83%	-3%		0%	0%	1%	0%	
	4%	9%	-3%	89%		1%	0%	0%	2%	
ρ_{ij}	100%	39%	20%	15%	+-	0%	0%	0%	0%	
	39%	100%	0%	15%		0%	0%	0%	0%	
	20%	0%	100%	0%		0%	0%	0%	0%	
	15%	15%	0%	100%		0%	0%	0%	0%	

The change in marginal distribution assumption has had a measurable but relatively small impact on the results. This is consistent with the theory in Appendix 3.

We now ask what happens if we make the dependency structure more conservative by using a Student-t copula with 10 degrees of freedom.

Not surprisingly, the results are more conservative. What is surprising, to the author, is that the aggregate result is still less than the starting point in Step 1. While research on the appropriate choice of copula is still in its early days (see Appendix 4), a number of published reports suggest that a t-copula with 10 degrees of freedom is a reasonable model of some real world dependency structures.

To close this section, we consider what happens when the copula degree of freedom parameter goes all the way down to three. This adds more tail dependency, so the result gets more conservative.

TABLE 2.5 Step 5

c_i	4	2.5	2	1.5					
ν_i	0.463	10	25	0.330		*Copula*	$\nu=$ 10		
	Lg Nrml	Stu-t	Stu-t	Pareto					
D	64.7%				+-	0.1%			
D_i	87%	65%	44%	34%	+-	0.4%	0.2%	0.4%	0.9%
D_{ij}	102%	37%	20%	12%	+-	1%	0%	0%	1%
	37%	94%	5%	15%		0%	0%	0%	0%
	20%	5%	96%	0%		0%	0%	1%	1%
	12%	15%	0%	89%		1%	0%	1%	2%
ρ_{ij}	100%	39%	20%	17%	+-	0%	0%	0%	0%
	39%	100%	0%	15%		0%	0%	0%	0%
	20%	0%	100%	0%		0%	0%	0%	0%
	17%	15%	0%	100%		0%	0%	0%	0%

TABLE 2.6 Step 6

c_i	4	2.5	2	1.5					
ν_i	0.463	10	25	0.330		*Copula*	$\nu =$ 3		
	Lg Nrml	Stu-t	Stu-t	Pareto					
D	70.5%				+-	0.3%			
D_i	87%	73%	55%	43%	+-	0.5%	0.5%	0.5%	0.6%
D_{ij}	100%	49%	26%	27%	+-	1%	1%	1%	1%
	49%	106%	9%	25%		1%	1%	1%	1%
	26%	9%	128%	3%		1%	1%	1%	1%
	27%	25%	3%	84%		1%	1%	1%	2%
ρ_{ij}	100%	39%	20%	22%	+-	0%	0%	0%	0%
	39%	100%	0%	15%		0%	0%	0%	0%
	20%	0%	100%	0%		0%	0%	0%	0%
	22%	15%	0%	100%		0%	0%	0%	0%

We now have a result that is more conservative than the starting point. Many practitioners would consider a t-copula with three degrees of freedom to be a fairly conservative dependency structure. This result suggests, to the author, that the diversification benefit that arises when aggregating risks with tail shape $\xi < 1/2$ can be a material issue.

What is interesting about this example is that we can now ask what would happen if the marginal distributions were all made more heavy-tailed by changing them to Student-t distributions with three degrees of freedom. The answer is that the result would go back down to the starting point ($D = 66.8\%$) because the model would become elliptical again. This somewhat counterintuitive result is an example of how elliptical symmetry can confound the otherwise simple intuition that heavier tails aggregate in a more conservative fashion.

Having gone through this modeling scenario, the company must now decide what to do. A debate now develops because no one really knows what the right copula is. One party argues that the result in Step 5 is appropriate because it is consistent with some known studies. Another party argues that, since we know so little about the true copula, the company should use the more conservative result in Step 6. The debate is finally brought to closure when someone remembers how uncertain the initial rank correlation matrix was. A simple alternative is proposed where a flat 10% is added to all of the starting rank correlations. The result is in Table 2.7.

TABLE 2.7

c_i	4	2.5	2	1.5
D	71.2%			
D_i	88%	72%	51%	51%
D_{ij}	100%	50%	30%	30%
	50%	100%	10%	30%
	30%	10%	100%	10%
	30%	30%	10%	100%

This simplified approach has produced a result very similar to Step 6 and is also much easier to explain to a non-technical audience. The working group concluded by recommending the simple model in Table 2.7. They reached a practical conclusion but felt comfortable that they had worked through the significant issues and could defend their position if challenged by other knowledgeable risk professionals. They were also aware that the entire issue has to be revisited as additional theoretical or practical insights become available.

3. Conclusions

The author believes there are two sets of conclusions that can be drawn from the analysis reported here.

1. It is possible to use fairly simple models for aggregating capital even though the real aggregation issues can be very complex. There are two reasons.

   - The theoretical errors made in assuming an elliptical model will tend to offset in many circumstances. Given this fact and the lack of precision in most model parameters, it is not clear that the apparent precision of a more complex model is real.

   - Even if a more complex model is justified, it can always be approximated locally by a tail correlation type formula. The tail correlation matrix, as defined in this paper, can be a useful tool for communicating results no matter what the model.

2. When building complex models, a large number of assumptions must be made about issues such as copulas and marginal distributions. We have seen that, at least for the marginal distributions, we don't need to know every detail about the distribution in order to get meaningful results as explained in Appendix 3. In terms of the copula assumption, it is not yet clear to the author which copula properties are crucial to the outcome and which are not. See Appendix 2 for more detail.

Acknowledgments

The work reported here represents the completion of a project started in the spring of 2003 while the author was an adjunct professor at the University of Waterloo. The author would like to acknowledge many useful conversations with Prof. H. Panjer and the contributions of J. Rodriguez and V. Vecchione who worked on an early version of the simulation software.

More recently, the author gratefully acknowledges the support of the Group Risk Department of AEGON NV in allowing this work to be completed. In particular, the simulation software used for the examples in this paper was written by M. Matkovic of AEGON NV.

Appendix 1. The Calculus of Risk Measures

In Section 1.4 a number of results were stated relating to the derivatives of some capital aggregation functions. The purpose of this appendix is to provide more detail on those results. The process we go through is to establish the results when a smooth distortion function is used as a risk measure and then consider what happens when we specialize to *CTE* or *VaR*.

A distortion function is a non decreasing function $g:[0,1]\to[0,1]$ such that $g(0)=0$ and $g(1)=1$. Given any such function and a random variable *X* we can calculate the risk measure:

$$R_g[X]=\int x dg[F(x)],$$

where $F(x)=\Pr(X\le x)$ is the distribution function of *X*. The mapping R_g is often called a distortion measure because the calculation is equivalent to calculating the mean of *X* with respect to a distorted probability function $F^*(x)=g[F(x)]$. Some well known examples of distortion measures are listed in the table below.

Name	Formula
Value at Risk (*VaR*)	$g(t)=\begin{cases}0 & t\le 1-\alpha\\ 1 & t>1-\alpha\end{cases}$
Conditional Tail Expectation (*CTE*)	$g(t)=\begin{cases}0 & t<\alpha\\ (t-\alpha)/(1-\alpha) & t\ge\alpha\end{cases}$
Wang Transform	$g(t)=\Phi[\Phi^{-1}(t)+\Phi^{-1}(\alpha)]$
Block Maxima	$g(t)=t^m$

If the distortion function is continuous and convex,[15] then the resulting risk measure is coherent in the sense of Artzner et al. In the table above, all of the risk measures are coherent except value at risk.

Now assume we have *n* risks $X_1,\ldots,X_n$ and *n* real numbers $e_1,\ldots,e_n$ and let $X=\sum_i e_i X_i$; then we can define a function:

$F(e_1,\ldots,e_n)=R_g(\sum_i e_i X_i)=R_g(X)$.

Given this notation, the main results of this appendix are

$$\frac{\partial F}{\partial e_i}=\int E[X_i\mid X=x]dg[F_X(x)]$$

[15] See, for example, Hurlimann, W., "Distortion Risk Measures and Economic Capital," *North American Actuarial Journal* (2004).

and $$\frac{\partial^2 F}{\partial e_i \partial e_j} = \int [Cov(X_i, X_j) \mid X = x] dg'[F_X(x)], \quad g'(t) = dg/dt.$$

When the distortion function g is smooth and the random variables $X_1,...,X_n$ have a smooth probability density function, the proof of the above result is an exercise in multi-variable calculus. We omit the details.

If the distributions or risk measures involved are not smooth, some additional care is required when deriving the above results. While we don't consider non-smooth probability distributions in this paper, both *VaR* and *CTE* are based on distortion functions that are not infinitely differentiable. However, it is not hard to check that if we have a sequence of smooth distortion functions $g_n(t)$ that converges, pointwise, to a limiting distortion function $g(t)$, then the formulas above continue to hold provided we interpret the integration operations in the sense of generalized functions.

For example, if the risk measure is *CTE*(*u*) we find:

$$dg[F(x)] = \begin{cases} 0 & F(x) < u \\ \dfrac{f(x)dx}{(1-u)} & F(x) \geq u \end{cases},$$

which can be used to show that $\frac{\partial F}{\partial e_i} = \int E[X_i \mid X = x] dg[F_X(x)] = E[X_i \mid X \geq VaR_X(u)]$.

For second derivatives the quantity $dg'[F(x)]$ must be interpreted as a generalized function so:

$$\begin{aligned} \frac{\partial^2 F}{\partial e_i \partial e_j} &= \int [Cov(X_i, X_j) \mid X = x] dg'[F_X(x)], \\ &= [Cov(X_i, X_j) \mid X = VaR_X(u)] \frac{f_X(VaR(u))}{1-u}. \end{aligned}$$

To get the results used earlier in the paper we must consider the relationship between the stand-alone capital $c_1,...,c_n$ and the exposure variables $e_1,...,e_n$. This relationship is clearly $c_i(e_i) = R_g(e_i X_i) = e_i R_g(X_i)$. The derivatives of the function *F* with respect to one of the capital variables is then:

$$\begin{aligned} \frac{\partial F}{\partial c_i} &= \frac{\partial F}{\partial e_i}\frac{\partial e_i}{\partial c_i} = E[X_i \mid X \geq VaR_X(u)] \frac{1}{R_g(X_i)}, \\ &= \frac{E[X_i \mid X \geq VaR_X(u)]}{c_i}, \quad when\ e_1 = ... = e_n = 1. \end{aligned}$$

The same argument shows that:

$$\frac{\partial^2 F}{\partial c_i \partial c_j} = [Cov(X_i, X_j) \mid X = VaR_X(u)] \frac{f_X(VaR(u))}{c_i c_j (1-u)}.$$

The issues associated with putting these results into practice are discussed in Appendix 4. The main challenge is estimating the conditional covariance $[Cov(X_i, X_j) \mid X = VaR_X(u)]$.

If one chooses to work with *VaR* as a risk measure, then we can invert the relation $CTE(u) = \frac{1}{1-u}\int_u^1 VaR(u)du$ to yield $VaR(u) = -\frac{d}{du}(1-u)CTE(u)$ hence

$$\begin{aligned}
\frac{\partial VaR}{\partial e_i} &= -\frac{\partial}{\partial e_i}\frac{d}{du}(1-u)CTE(u) \\
&= -\frac{d}{du}(1-u)\frac{\partial}{\partial e_i}CTE(u) \\
&= E[X_i \mid X = VaR_X(u)]
\end{aligned}$$

and

$$\frac{\partial^2 VaR}{\partial e_i \partial e_j} = -\frac{d}{du}\{Cov[(X_i, X_j) \mid X = VaR_X(u)] f_X(VaR_X(u)\}.$$

One implication of these results is that the technical issues encountered when working with *VaR* are more challenging than when working with *CTE*. From a purely technical perspective, *CTE* is probably the easiest risk measure with which to work.

A second implication is that we cannot guarantee that the tail correlation matrix for *VaR* will be positive semi-definite.

We can state that the tail correlation for *CTE* is positive semi-definite because the tail correlation is the sum of two components, both of which are semi-definite.

$$D_{ij} = \frac{\partial C}{\partial c_i}\frac{\partial C}{\partial c_j} + C[Cov(X_i, X_j) \mid X = VaR_X(u)]\frac{f_X(VaR(u))}{(1-u)c_i c_j}.$$

In the general situation we can write the tail correlation matrix as:

$$D_{ij} = \frac{\partial C}{\partial c_i}\frac{\partial C}{\partial c_j} + \frac{C}{c_i c_j}\int [Cov(X_i, X_j) \mid X = x] dg'[F_X(x)].$$

The second term above is a linear combination of positive semi-definite matrices with coefficients given by $dg'[F(x)] = g''[F(x)]dF(x)$. So if the distortion function is convex, i.e., $g''[F(x)] \geq 0$, we can guarantee that the tail correlation matrix will be positive semi-definite. This is the basis for the claim made in Section 1 that coherent risk measures have positive semi-definite tail correlation matrices.

Appendix 2. Choosing the Copula and Related Parameters

Despite a large academic literature on the subject of copulas,[16] the author does not believe there is a comprehensive professional consensus as to which copulas should or should not be used for risk aggregation modeling. Early workers in the credit risk field started by using Gaussian copulas due to their mathematical simplicity. These models were then criticized, appropriately, for their lack of tail dependence.

The Student-t copula is used in this paper because it is easy to work with, it exhibits tail dependence, and the parameters are fairly easy to understand. These reasons do not mean it is the most appropriate copula for the risk aggregation application.

Many other copula families have been described in the literature but it is not clear, to the author, why any of them is necessarily doing a better job than the Student-t copula. Ideally, one would like a result that allows practitioners to understand which copula properties are crucial to the aggregation issue and which properties can be safely ignored. Until that insight is available, the author believes it makes sense to stick with something simple.

Two studies which suggest that a Student-t copula with 10 degrees of freedom is a reasonable model for the correlation of returns between some asset classes are Mashal, Naldi and Zeevi[17] and Kiole, Koedijk and Verbeek.[18] The last study compares the relative ability of the Gaussian, Student-t and Gumbel copulas to explain the dependency structure of equity, fixed income and real estate returns and concludes that the Student-t copula does a better job.

[16] See MFE Chapters 5 and 7 for a very good overview.
[17] Mashal, R., Naldi, M., and Zeevi, A., "On the Dependence of Equity and Asset Returns," *Risk* Magazine (October 2003).
[18] Kole,E., Koedijk, K., and Verbeek, M., "Selecting Copulas for Risk Management," Preprint (September 2006).

Appendix 3. Choosing the Marginal Distribution and Related Parameters

This appendix documents some ideas that are needed to understand the paper's claim that we don't need to know every detail about the marginal distributions used in our simulation models. The appendix starts by developing some technical results related to extreme value theory and ends by surveying the tail behavior of some standard risk models.

Some Technical Results

Let X be a real valued random variable and set $F(x) = \Pr(X \le x)$ and define $VaR(u) = F^{-1}(u) = \inf\{x \mid F(x) \ge u\}$. The two functions $CTE(u)$ and $CTV(u)$ are defined for $0 \le u \le 1$ by:

$$CTE(u) = \frac{1}{1-u}\int_u^1 F^{-1}(v)dv\,,\quad CTV(u) = \frac{1}{1-u}\int_u^1 [F^{-1}(v)]^2\,dv - CTE(u)^2.$$

If the random variable is continuous, then $CTE(u)$ is just the conditional tail expectation or *Tail VaR* of X , $CTE(u) = E[X \mid X \ge VaR(u)]$. Similarly, $CTV(u)$ is the conditional tail variance when it exists.

For any risk with finite variance, we can define a tail shape function $\xi(u)$ by the equation:

$$1 - 2\xi(u) = \frac{[CTE(u) - VaR(u)]^2}{CTV(u)}.$$

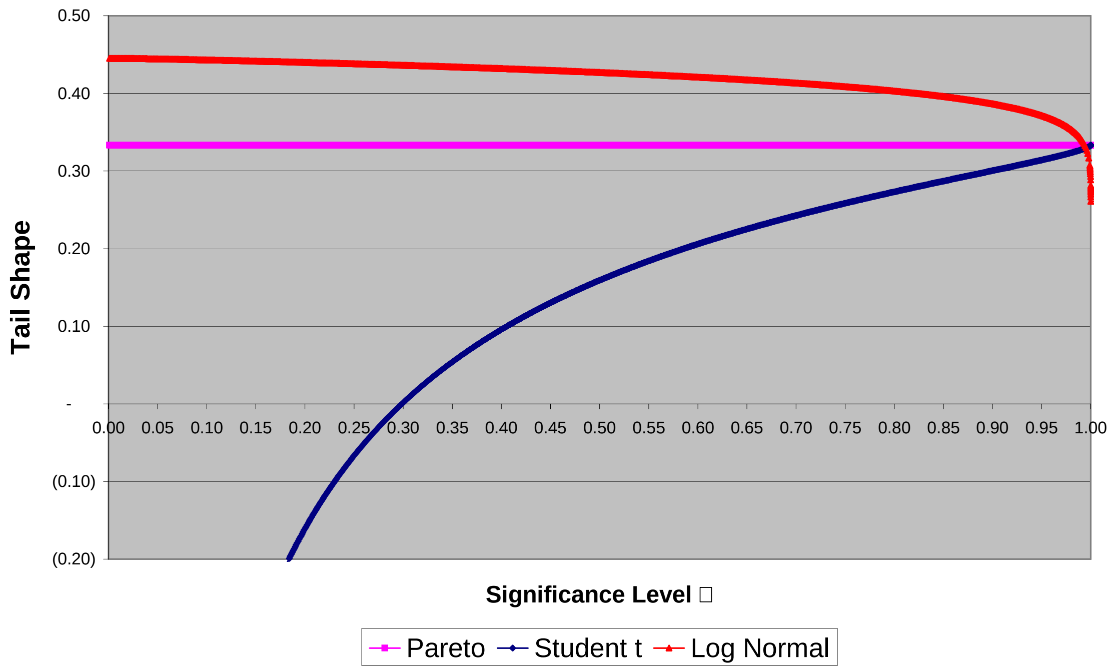
There are many reasons why this is a useful concept. The two most important are:

- The tail shape is independent of any location/scale parameters. Both the numerator and denominator are independent of location, so, by taking the ratio, we get something independent of both location and scale, i.e., a shape measure.

- For most, though not all, distributions of interest to risk managers, the tail shape function is defined and reasonably well behaved for all $0 \le u \le 1$. Note that $-\infty < \xi \le 1/2$ for all *u*.

The graph below shows the behavior of the tail shape function for three well known probability distributions.

TAIL SHAPE FUNCTION FOR 3 MARGINAL MODELS



The Pareto distribution is defined by the formula $F^{-1}(u) = A + B(1-u)^{-\hat{\xi}}$ where $A, B, \hat{\xi} \neq 0$ are all constants such that $B\hat{\xi} > 0$. It is easy to verify that if $\hat{\xi} < 1/2$ then the tail shape is well defined and $\xi(u) = \hat{\xi}$ for all *u*, which explains the choice of notation. If $\hat{\xi} \geq 1/2$ then *CTV* does not exist and if $\hat{\xi} \geq 1$ *CTE* does not exist.

The Student-t example here has *n* = 3 degrees of freedom. The tail shape function starts out at $-\infty$ and then grows monotonically until it reaches the limiting value 1/*n*.

The lognormal model is defined by $F^{-1}(u) = A + B\exp[\sigma\Phi^{-1}(u)]$. The parameter σ was chosen so that the tail shape of the lognormal model matches that of the Student-t model when $u = .995$. Tail shape is positive for all $u < 1$ with limiting value of 0 as $u \to 1$.

Risks are usually referred to as heavy-tailed if $\xi(u) > 0$ in a neighborhood of $u = 1$ and light-tailed if $\xi(u) \leq 0$ in a neighborhood of $u = 1$. By that standard all of these examples are heavy-tailed.

Examples of light-tailed risks are the exponential distribution $F^{-1}(u) = A + B\ln(1-u)$ for which $\xi(u) = 0\ \forall u \in [0,1]$ and the normal distribution which is the $n \to \infty$ limit of the Student-t.

We now prove a result that basically says most continuous distributions are locally Pareto in the sense that if $u^* \in [0,1]$ and u is in a sufficiently small neighborhood around u^*, then there are constants $A(u^*), B(u^*)$ such that $F^{-1}(u) \approx A(u^*) + B(u^*)(1-u)^{-\xi(u^*)}$. Letting $u^* \to 1$, this is the core result of Extreme Value Theory although it is not usually presented in this way. The derivation given here is much simpler than the one used in most textbooks because they usually do not assume the *CTV* exists.

Start by doing a little bit of calculus. If we differentiate the expressions for *CTE*(u) and *CTV*(u) we find:

$$\frac{dCTE(u)}{du} = \frac{CTE(u) - VaR(u)}{1-u},$$
$$\frac{dCTV(u)}{du} = CTV(u)\frac{2\xi(u)}{1-u}.$$

Note that the last expression tells us that a risk is heavy/light-tailed according to whether *CTV*(u) is increasing or decreasing in the tail.

Integrating the last expression from u^* to u we find:

$$CTV(u) = CTV(u^*)\exp[\int_{u^*}^{u}\frac{2\xi(v)}{1-v}dv],$$

and if we confine our attention to u in a small enough neighborhood of u^* we can assume $\xi(u) \approx \xi(u^*)$ in the integral and conclude:

$$CTV(u) \approx CTV(u^*)\left(\frac{1-u}{1-u^*}\right)^{-2\xi(u^*)}.$$

Now go back to the derivative of *CTE*(u) and use the definition of tail shape and the result above to write:

$$\begin{aligned}\frac{dCTE(u)}{du} &= \frac{CTE(u) - VaR(u)}{1-u}, \\ &= \frac{\sqrt{CTV(u)(1-2\xi(u))}}{1-u}, \\ &= \sqrt{CTV(u^*)}\exp[\int_{u^*}^{u}\frac{\xi(v)}{1-v}dv]\frac{\sqrt{1-2\xi(u)}}{1-u}.\end{aligned}$$

Integrating this expression from u^* to u we find:

$$CTE(u) = CTE(u^*) + \sqrt{CTV(u^*)}\int_{u^*}^{u} \exp[\int_{u^*}^{w} \frac{\xi(v)}{1-v} dv] \frac{\sqrt{1-2\xi(w)}}{1-w} dw.$$

Again assuming $\xi(u) \approx \xi(u^*) \neq 0$ we find:

$$\begin{aligned}
CTE(u) &\approx CTE(u^*) + \sqrt{CTV(u^*)}\int_{u^*}^{u} \left(\frac{1-w}{1-u^*}\right)^{-\xi(u^*)} \frac{\sqrt{1-2\xi(u^*)}}{1-w} dw, \\
&= CTE(u^*) + \frac{\sqrt{CTV(u^*)}\sqrt{1-2\xi(u^*)}}{(1-u^*)^{-\xi(u^*)}} \int_{u^*}^{u} (1-w)^{-\xi(u^*)-1} dw, \\
&= CTE(u^*) + \frac{\sqrt{CTV(u^*)}\sqrt{1-2\xi(u^*)}}{\xi(u^*)} \left[\left(\frac{1-u}{1-u^*}\right)^{-\xi(u^*)} - 1\right].
\end{aligned}$$

If $\xi(u) \approx \xi(u^*) = 0$ we get the modified result:

$$CTE(u) \approx CTE(u^*) + \sqrt{CTV(u^*)} \ln\left(\frac{1-u}{1-u^*}\right).$$

Finally, we can get an expression for the *VaR* by using:

$$\begin{aligned}
F^{-1}(u) &= VaR(u) \\
&= CTE(u) - \sqrt{CTV(u)[1-2\xi(u)]} \\
&= CTE(u^*) + \sqrt{CTV(u^*)}\int_{u^*}^{u} \exp[\int_{u^*}^{w} \frac{\xi(v)}{1-v} dv] \frac{\sqrt{1-2\xi(w)}}{1-w} dw \\
&\quad - \sqrt{CTV(u^*)} \exp[\int_{u^*}^{u} \frac{\xi(v)}{1-v} dv] \sqrt{[1-2\xi(u)]}
\end{aligned}$$

When $u^* = 0$ we have $CTE(0) = \mu$ and $\sqrt{CTV(0)} = \sigma$, and we get the very general expression:

$$F^{-1}(u) = \mu + \sigma\left(\int_0^u \frac{\sqrt{1-2\xi(w)}}{1-w} \exp(\int_0^w \frac{\xi(v)}{1-v} dv) dw - \sqrt{1-2\xi(u)} \exp(\int_0^u \frac{\xi(w)}{1-w} dw)\right),$$

which shows that, if we know the mean, standard deviation and tail shape $\xi(u)$ for all values of *u*, we can recover the entire distribution.[19]

Returning to the situation where the tail shape is locally constant, we find:

$$VaR(u) \approx CTE(u^*) + \frac{\sqrt{CTV(u^*)}\sqrt{1-2\xi(u^*)}}{\xi(u^*)}\left[\left(\frac{1-u}{1-u^*}\right)^{-\xi(u^*)} - 1\right] - \sqrt{CTV(u^*)}\left(\frac{1-u}{1-u^*}\right)^{-\xi(u^*)}\sqrt{1-2\xi(u^*)}$$

which clearly has the form $F^{-1}(u) \approx A(u^*) + B(u^*)(1-u)^{-\xi(u^*)}$ as originally claimed.

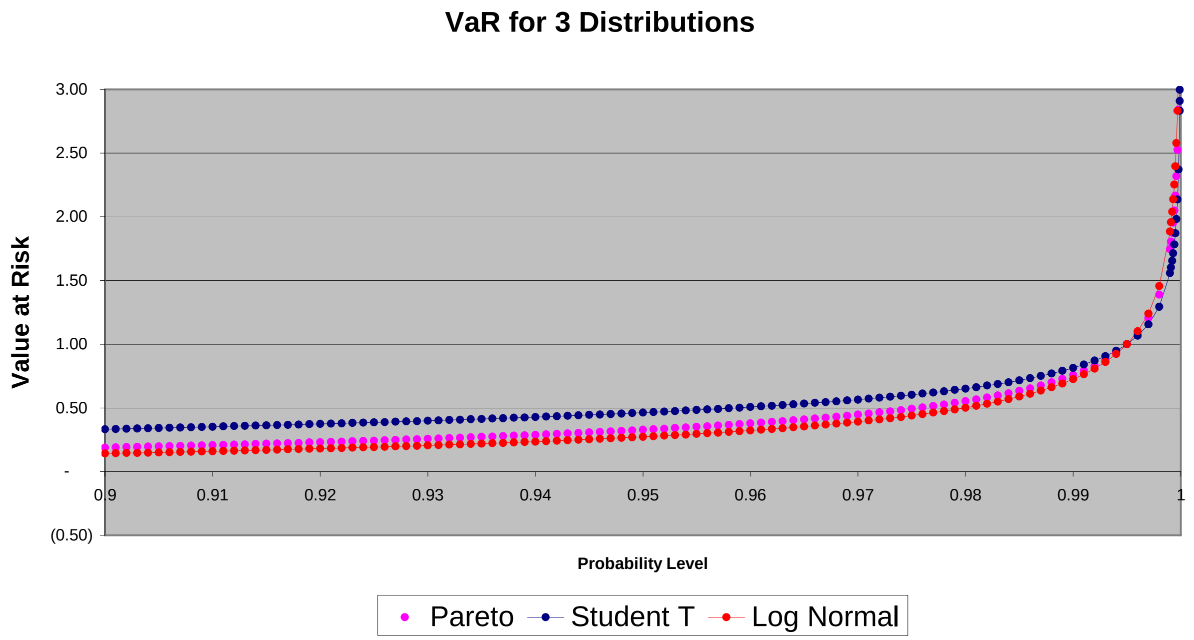
A short summary of the preceding discussion is that we can always approximate a *VaR* function locally by a Pareto distribution. However, on the same assumptions, that means we can also use other models as an approximation. It merely remains to be seen which "proxy" is the most useful. In particular we can draw the following conclusion:

> *Two distributions which are calibrated to have the same location/scale parameters and have the same tail shape at the point* $u = u^*$ *will have similar values for* $VaR(u)$ *for u in a neighborhood of* u^*.

The chart below gives visual evidence of the above result. The chart plots the *VaR* function above *u* = .90 for a lognormal, Pareto and Student-t risk. The parameters of each model have been set so that

- The mean of the distribution is zero (location)
- The *CTE*(.99) values are equal (scale)
- The remaining parameter was calibrated to match the tail shape, $\xi(.995)$, of a Student-t model with three degrees of freedom.

[19] This does not mean that we can specify $\xi(u)$ arbitrarily. The fact that $dF^{-1}(u)/du \geq 0$ implies that $\eta(u) = \dfrac{1-\xi(u)}{(1-2\xi(u))(1-u)}$ is an increasing function of *u*.



When $u < .995$, there is a clear ordering among the VaR functions with the Student-t being the highest and the lognormal value being the lowest. Those relationships all reverse once u exceeds a point slightly higher than .995.

Sensitivity testing shows that, when aggregating capital using *CTE*(.99), it is conservative to use the Student-t model above relative to the other two choices.[20] This suggests that it is the region below .995 that is driving the result. A high level theoretical argument that helps to explain this result is to consider the following set of inequalities for a first order diversification factor D_i. We have:

$$\begin{aligned} D_i &= \frac{E[X_i \mid X \geq VaR(.99)]}{c_i}, \\ &\approx \frac{E[X_i \mid X = VaR(.995)]}{c_i} \\ &\leq \frac{VaR(X_i,.995)]}{c_i} \approx 1 \end{aligned}$$

The point is that it is the values of X_i below its .995 *VaR* level that are having the most impact on the diversification factor.

[20] See Step 4 in the modeling scenario in Section 2.

Tail Shape for Some Standard Risks

Having argued that $\xi(.995)$ is an important parameter to get right, it is appropriate to ask how to set this parameter. Below are some short notes on the four broad risk categories used in the examples.

Operational Risk. In 2003 and 2004, the Federal Reserve Bank of Boston published studies[21,22] of operational risk losses. The 2003 study looked at the data available in public databases and concluded that operational risk is either heavy-tailed or very heavy-tailed. The 2004 study used data from six large U.S. banks that is not in the public domain. The same conclusion was reached. The value $\xi(.995) = .033$ used in this paper's examples is probably at the low end of the range of reasonable tail shape values.

Credit Risk. To get some insight into credit risk issues, one can look at Oldrich Vasicek's model for credit risk contagion.[23] This model is intended to estimate the default rate for a portfolio of bonds assuming Merton's structural model where a bond defaults if the market value of the issuer drops below the debt amount. There are three key assumptions:

1. The portfolio is large enough and diverse enough that the law of large numbers can be applied.
2. The stand-alone probability of default for any one bond is q.
3. The stock of each issuer follows a lognormal process, and the correlation between all pairs of returns is ρ.

For $0 \le u \le 1$ let $Q(u)$ be the portfolio probability of default at the level *u*, i.e., the probability of the portfolio loss being less than a fraction $Q(u)$ of the total is just *u*. In a 2002 *Risk* Magazine paper, Vasicek showed that the simplifying assumptions above imply:

$$Q(u) = \Phi\left[\frac{\sqrt{\rho}\,\Phi^{-1}(u) + \Phi^{-1}(q)}{\sqrt{1-\rho}}\right], \Phi = \text{cumulative normal}.$$

While this may not be the most intuitive-looking formula, it is simple enough to work with, and it does capture some key risk issues. For example, if the expected default level is $q = .02$ and $\rho = .10$, the .995 loss level is $Q(.995) = .0957$ of the portfolio.

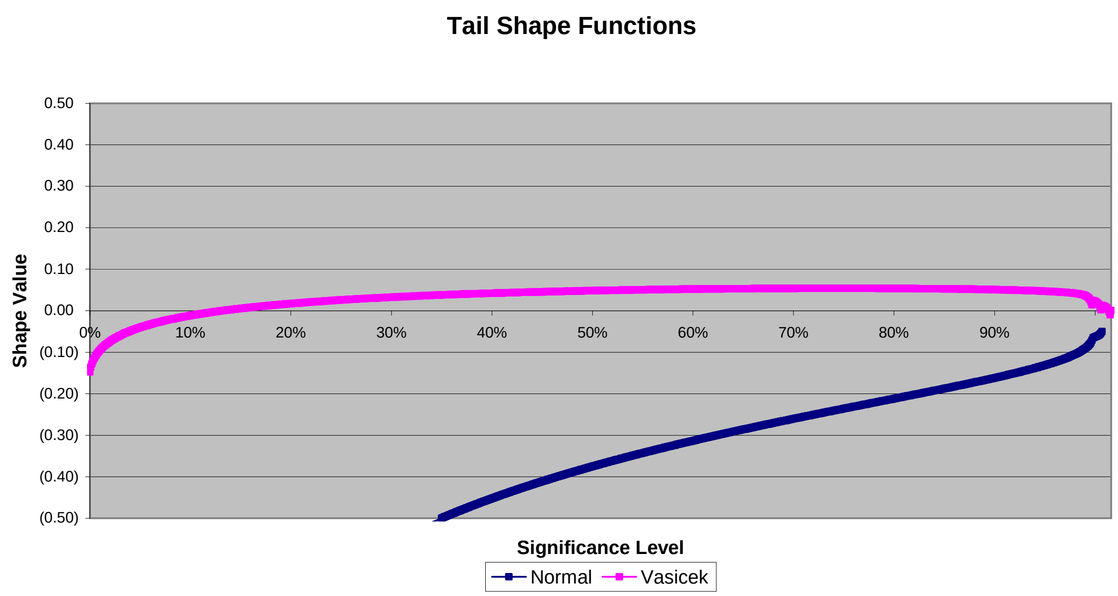
The chart below shows the general behavior of the tail shape function for the Vasicek model. For many choices of parameters, the tail shape starts out negative, rises to a positive maximum and then declines to zero. For comparison, the chart also shows the tail shape for the normal distribution which also has a limiting value of zero.

---

21 De Fontnouvelle, P., DeJesus-Rueff, V., and Jordan, J., "Capital and Risk: New Evidence on Implications of Large Operational Losses." 2003.

22, De Fontnouvelle, P., Rosengren, E., and Jordan, J., "Implications of Alternative Operational Risk Modeling Techniques." 2004.

23 Vasicek, O., "Loan Portfolio Value." *Risk* Magazine (2002), reprinted July 2007.



The tables below show the tail shape for this model at both the 99.0% and 99.5% level for a range of input parameters.

Tail Shape @ 99.0% for Vasicek's Credit Loss Model

		q: 0.1%	q: 0.5%	q: 1.0%	q: 2.0%
rho	0.5%	(0.02)	(0.03)	(0.04)	(0.04)
rho	1.0%	0.01	(0.01)	(0.02)	(0.03)
rho	2.0%	0.04	0.01	0.00	(0.01)
rho	5.0%	0.09	0.05	0.04	0.02
rho	10.0%	0.14	0.09	0.07	0.04
rho	20.0%	0.20	0.12	0.09	0.04

Tail Shape @ 99.5% for Vasicek's Credit Loss Model

		q: 0.1%	q: 0.5%	q: 1.0%	q: 2.0%
rho	0.5%	(0.01)	(0.03)	(0.03)	(0.04)
rho	1.0%	0.01	(0.01)	(0.02)	(0.02)
rho	2.0%	0.04	0.01	0.00	(0.01)
rho	5.0%	0.08	0.05	0.03	0.02
rho	10.0%	0.13	0.08	0.06	0.03
rho	20.0%	0.18	0.10	0.07	0.03

The first thing to notice is that the tail gets heavier as the risks become more correlated. This is clearly reasonable. More interestingly, we also see that the tail shape gets smaller as the probability of default increases. This is similar to the behavior of risks for options, which get lighter as the option moves into the money.

Based on this evidence one could consider credit risk to be a light- to medium-tailed risk depending on the credit quality of the portfolio.

Mismatch Risk. If we start with the presumption that mismatch risk is fundamentally being driven by a lognormal model, then a reasonable place to start is the table below. It shows that it takes a huge volatility parameter to create even a medium-tailed risk.

$\blacklozenge^2$	90.0%	95.0%	99.0%	99.5%	99.9%	100.0%
230%	0.39	0.37	0.34	0.33	0.31	0.00
129%	0.28	0.27	0.25	0.24	0.22	0.00
84%	0.21	0.20	0.19	0.18	0.17	0.00
59%	0.16	0.15	0.15	0.14	0.13	0.00
44%	0.12	0.12	0.11	0.11	0.11	0.00
34%	0.09	0.09	0.09	0.09	0.09	0.00
27%	0.06	0.07	0.07	0.07	0.07	0.00
22%	0.04	0.05	0.05	0.06	0.06	0.00
10%	(0.03)	(0.01)	0.01	0.01	0.02	0.00
5%	(0.06)	(0.04)	(0.02)	(0.01)	(0.00)	0.00
3%	(0.08)	(0.06)	(0.03)	(0.03)	(0.02)	0.00
2%	(0.09)	(0.07)	(0.04)	(0.04)	(0.02)	0.00
2%	(0.10)	(0.08)	(0.05)	(0.04)	(0.03)	0.00
1%	(0.11)	(0.09)	(0.06)	(0.05)	(0.03)	0.00
1%	(0.12)	(0.10)	(0.06)	(0.05)	(0.04)	0.00
0%	(0.14)	(0.11)	(0.08)	(0.07)	(0.05)	0.00

In this table, the leftmost column has the value of the parameter σ^2 that makes the tail shape at $u = .995$ agree with a Student-t model. The top row corresponds to *n*=3. The lognormal model is clearly capable of exhibiting heavy-tailed behavior even though the limiting value of the tail shape function is always 0.

This is clearly not a definitive argument, but it does suggest that, in the absence of special issues, mismatch risk should not be a source of risk with tail shape > ½.

Underwriting Risk. There are known examples of underwriting risk that have the potential to be very heavy-tailed, though very few of them occur in the life insurance industry. One possible life insurance example would be a contagion event such as a repeat of the 1918 flu epidemic scenario. However, more in-depth modeling suggests this risk is qualitatively similar to Vasicek's credit risk model described earlier since they are both basically contagion models.

The arguments presented here are intended to be indicative rather than comprehensive. The point is that very heavy-tailed risks are the exception rather than the rule, so we would expect aggregation approaches based on elliptical models to have an element of conservatism relative to more sophisticated models.

Appendix 4. Estimating Tail Correlation Using the Direct Method

It was argued in Section 1 of this paper that it was reasonable to assume that the capital aggregation function is homogeneous of degree 1 in its arguments, meaning that $\lambda > 0 \Rightarrow C(\lambda c_1,..., \lambda c_n) = \lambda C(c_1,..., c_n)$. Any function with this property satisfies two conditions that follow from differentiating the above equation with respect to λ and then setting $\lambda = 1$.

Differentiating once we find:

$$C = \sum_i \frac{\partial C}{\partial c_i} c_i \ ,$$

and differentiating a second time we get the result:

$$\sum_j \frac{\partial^2 C}{\partial c_i \partial c_j} c_j = 0.$$

The first result can be used to show that if $D_i = \partial C / \partial c_i$ then the approximation $C \approx \sum_i D_i (c_k{}^0) c_i$ is the same as a first order Taylor expansion about the point c_i^0.

If, in addition, the function C is positive and if we put $D_{ij} = \dfrac{1}{2} \dfrac{\partial^2 C^2}{\partial c_i \partial c_j}$, then the approximation $C \approx \sqrt{\sum_{i,j} D_{ij} (c_k{}^0) c_i c_j}$ has second order contact with the function C in a neighborhood of the base point c_i^0.

When there is no closed form expression telling us how to aggregate capital we need tools that can estimate the quantities described above from real or simulated data. In general, the required formulae vary according to the choice of risk measure. For the case where the risk measure is *CTE*(*u*) and $X = \sum X_i$, we have argued in Appendix 1 that:

$$\frac{\partial C}{\partial c_i} = \frac{E[X_i \mid X \geq X_u]}{c_i},$$

$$\frac{\partial^2 C}{\partial c_i \partial c_j} = \frac{Cov[(X_i, X_j) \mid X = X_u]}{(1-u) c_i c_j} f_X(X_u),$$

where X_u is the *VaR* of *X* at the level *u* and $f_X(x)$ is the probability density function of *X*.

The tail correlation matrix is then given by:

$$D_{ij} = \frac{1}{2}\frac{\partial^2 C^2}{\partial c_i \partial c_j} = \frac{\partial C}{\partial c_i}\frac{\partial C}{\partial c_j} + C\frac{\partial^2 C}{\partial c_i \partial c_j}.$$

Estimating first derivatives is not problematic; the challenge is estimating the second derivatives. There are two general methods for doing this

- Direct Methods that try to estimate D_{ij} using the formula above. A random sample of values is used to estimate X_u, and then adjust the data to get a sample from the conditional distribution given $X = X_u$. We will show that this can be done fairly easily when there are practical ways of calculating the probability density of the copula.

- Indirect Methods. The idea here is to use the simulation model or data to estimate the risk measure at a number of pivot points c_i^A in a neighborhood of the base point c_i^0. If $C^A = C(c_1^A,...,c_n^A)$, then we should be able to estimate the tail correlation matrix by fitting a formula approximation of the form $C^A = \sqrt{\sum_{i,j} D_{ij} c_i^A c_j^A}$ to the empirical results.

The Indirect Method has several advantages over the Direct Method. These are

1. It requires no special formulae. It will work for any risk measure or copula, etc., although working with *CTE* makes it much easier to estimate the sampling covariance matrix $\Sigma^{AB} = Cov(C^A, C^B)$.

2. By choosing the pivot points appropriately, we can know in advance how far we can go from the base point and still have the approximation hold.

The main disadvantage of the Indirect Method is that it uses more computer resources than the Direct Method. Whether this is important or not will depend on the software platform being used.

Assume our simulation model has produced raw output consisting of an $N \times (n+2)$ array where the first n columns are samples from the component risks $X_{Ai}, i = 1,...,n$, the n+1'st column is the total risk $X_A = \sum_i X_{Ai}$, and the last column is a statistical weight factor W_A that would arise if an importance sampling[24] technique were being used. We assume the matrix has been sorted on the total risk so that $X_1 \geq X_2 \geq ...$ etc.

[24] The examples reported earlier in this paper did not use importance sampling so $W_A = 1/N$.

Let u be the confidence level that we are using to estimate *CTE* e.g. $u = .99$. Define the indicator variable H_A by:

$$H_A = \begin{cases} 1 & if \ \sum_{B=1}^{A} W_B \le 1-\alpha \\ 0 & otherwise \end{cases}.$$

Also let $\hat{A} = \inf\{A \mid H_A > 0\}$ then estimators for C, D_i and $VaR(X)$ are given by:

$$\hat{C} = \sum_{B=1}^{\hat{A}} X_B W_B \Big/ \sum_{B=1}^{\hat{A}} W_B,$$

$$\hat{D}_i = \frac{1}{c_i} \sum_{B=1}^{\hat{A}} X_{Bi} W_B \Big/ \sum_{B=1}^{\hat{A}} W_B.$$

$$V\hat{a}R(X) = X_{\hat{A}}$$

For finite N the C, D_i estimators are known to be negatively biased but the bias is typically much smaller than the sampling error.

In order to estimate tail correlation we need to estimate the second derivatives of the capital function from:

$$\frac{\partial^2 C}{\partial c_i \partial c_j} = \frac{Cov[(X_i, X_j) \mid X = X_u]}{(1-u)c_i c_j} f_X(X_u).$$

At a first glance this seems like an almost impossible task, but it can be solved with a mathematical trick. The trick is to use the available data to generate a sample from the conditional distribution of the component risks given that the total is at the *VaR* level.

Given an estimator $X_{\hat{A}}$ for $VaR(X)$, we can create a sample Y_{Ai} from the conditional distribution by setting:

$$Y_{Ai} = \begin{cases} X_{Ai} & i = 1,\ldots,n-1 \\ X_{\hat{A}} - \sum_{j=1}^{n-1} X_{Aj}, & i = n \end{cases}.$$

There is clearly nothing special about using the *n*'th risk as the balancing item here.

The problem with this sample is that the probability of observing it is $f(y_1,..,y_{n-1})$ and not $f(y_1,..,y_n)$ where $f(y_1,..,y_n)$ is the probability density function of our simulation model.

The solution is to introduce a statistical weight, which is the ratio of these two probabilities. Thus we define the weight:

$$J_A = \frac{f(y_{A1},...,y_{An})}{f(y_{A1},...,y_{A(n-1)})}.$$

If we can get our hands on this quantity, then estimators for $Y_i = E[X_i \mid X = VaR]$, $Y_{ij} = E[X_i X_j \mid X = VaR]$ and $f_X(VaR)$ can be calculated as follows:

$$\hat{Y}_i = (\sum_A Y_{Ai} W_A J_A) / \sum_A W_A J_A,$$
$$\hat{Y}_{ij} = (\sum_A Y_{Ai} Y_{Aj} W_A J_A) / \sum_A W_A J_A,$$
$$\hat{f}_X = (\sum_A W_A J_A) / \sum_A W_A.$$

The estimator for the second derivative term is then just

$$\hat{C}_{ij} = \frac{[\hat{Y}_{ij} - \hat{Y}_i \hat{Y}_j] \hat{f}_X}{c_i c_j (1-\alpha)}.$$

To calculate J_A, assume the simulation model's distribution function $F(x_1,...,x_n)$ is defined in terms of a second distribution function $G(t_1,...,t_n)$ by:

$$F(x_1,...,x_n) = G(G_1^{-1}(F_1(x_1)),...,G_n^{-1}(F_n(x_n))).$$

Here $G_i(t_i)$ and $F_i(x_i)$ are the marginal distribution functions. In the case of the t-copula, $G(t_1,...,t_n)$ is the multivariate Student-t distribution function and the G_i are the univariate marginal distribution functions . Differentiating the above equation once with respect to each variable yields:

$$f(x_1,...,x_n) = g(t_1,...,t_n) \frac{f_1(x_1)...f_n(x_n)}{g_1(t_1)...g_n(t_n)}, \quad G_i(t_i) = F_i(x_i)$$

where:

$$g(t_1,...,t_n) = \frac{\partial^n}{\partial t_1 ... \partial t_n} G(t_1,...,t_n), \qquad g_i(t_i) = \frac{d}{dt_i} G_i(t_i).$$

It follows then that the ratio we are looking for is:

$$J(y_1,...,y_n) = \frac{g(t_1,...,t_n)}{g(t_1,...,t_{n-1})} \frac{f_n(y_n)}{g_n(t_n)}, \quad G_i(t_i) = F_i(y_i).$$

If the model we have chosen is the t-copula, then all of the probability densities are known in closed form and can be easily programmed.

Based on all of the above work we now have an estimator for the tail correlation matrix:

$$\hat{D}_{ij} = \hat{D}_i \hat{D}_j + \hat{C} \frac{[\hat{Y}_{ij} - \hat{Y}_i \hat{Y}_j] \hat{f}_X}{c_i c_j (1-\alpha)}.$$

The method outlined above was first piloted in a spreadsheet and then rewritten in another language to get the results reported earlier in the paper. The fact that tail correlation is equal to linear correlation for elliptical models provides an important set of test cases that can be used to debug the software’s implementation.

Appendix 5. Estimating Tail Correlation Using the Indirect Method

The idea behind the Indirect Method is to estimate the aggregate capital at a number of points in a neighborhood of the base point and then fit a formula to the resulting estimates.

To fix notation, let c_i^0 denote the capital requirements at the base point; let c_i^A A=1,…,M denote the capital mix at a set of neighboring points; and let $\hat{C}^A = \hat{C}(c_1^A,...,c_n^A)$ be the estimated aggregate capital based on a set of simulations. We want to fit a formula of the form $\hat{C}^A \approx \sqrt{\sum_{i,j} D_{ij} c_i^A c_j^A}$ to the estimated capital amounts.

Here are some of the considerations for choosing the pivot points c_i^A.

1. Since the tail correlation matrix is symmetric, we need to estimate n diagonal elements and $\frac{n(n-1)}{2}$ off diagonal elements for a total of $\frac{n(n+1)}{2}$ unknowns. This is the minimum number of points that must be used and shows that, as the number of risks n in the model grows, the computational cost grows as n^2. For the Direct Method, the computational cost is of order n.

2. The pivot points should cover a neighborhood of the base point in a fairly even way. No particular direction should get special treatment.

3. There is sampling error in each of the estimates. We need to take this into account either when fitting a formula or assessing the goodness of fit afterwards. If the errors in the formula are comparable to the sampling error for a suitably large sample size N, then we should be happy with the result.

One approach is to create pivot points by rotating the base point c_i^0 in each of the $\frac{n(n-1)}{2}$ coordinate planes through the angles $\theta = \pm.10\frac{\pi}{2}$. The factor of $.10$ determines the size of the neighborhood we are using. This approach recognizes the scaling symmetry $C(\lambda c_1,...,\lambda c_n) = \lambda C(c_1,...,c_n)$ and treats each of the inputs c_i^0 in a symmetric way. The total number of pivot points is then $1 + 2 \times n(n-1)/2 = n^2 - n + 1 = M$. This exceeds the minimum as long as $n \geq 3$. This is clearly not the only way to choose pivot points.

The method of influence functions can be used to estimate the sampling error in each of these quantities.[25] Let $IF_{(C)}^A$ $A = 1,...,M$; $(C) = 1,...N$ be the (unsorted) influence function for the estimator $\hat{C}^A$.

---

[25] See Section 5 of the paper, "Variance of the CTE Estimator," by Manistre & Hancock, *North American Actuarial Journal* 9(2) (2005).

Then:

$$\Sigma^{AB} = \frac{1}{N^2}\sum_{(C)=1}^{N} IF_{(C)}^{A} IF_{(C)}^{B}$$

is an estimate of the sampling covariance $Cov(\hat{C}^A, \hat{C}^b)$. Let $\Sigma_{AB} = (\Sigma^{AB})^{-1}$ and let $F^A = \sqrt{\sum_{i,j} D_{ij} c_i^A c_j^A}$ be a formula value based on an estimated tail correlation matrix D_{ij}. If F^A is a reasonable estimate of the true aggregate capital C^A, and the sample size is large enough, then the quantity:

$$\chi^2 = \sum_{A,B} \Sigma_{AB}(F^A - \hat{C}^A)(F^B - \hat{C}^B)$$

has an approximate Chi-square distribution with M degrees of freedom. This suggests the following strategy:

1. Let G_{AB} be any positive definite symmetric matrix.

2. Estimate the tail correlation matrix by choosing D_{ij} to be the positive semi-definite matrix that minimizes the fit measure:

$$\sum_{A,B} G_{AB}(F^A - \hat{C}^A)(F^B - \hat{C}^B).$$

3. Decide on the quality of the fit by looking at the χ^2 statistic or by examining other measures such as relative absolute error, etc.

One possible choice for the fit measure is χ^2 itself. However, the optimization process will exploit any quirks in the estimated covariance matrix so this is not recommended.

Another choice is $G_{AB} = I_{AB}$, the identity matrix. Test work so far suggests that this leads to reasonable results.

Here is a short list of the pros and cons of the Direct Method versus the Indirect Method

- In the a spreadsheet environment, with $n = 4$, one iteration of the Indirect Method takes almost 10 times longer to run than one iteration of the Direct Method.

- The two methods produce the same result, within sampling error, but the precision of the Indirect Method does not seem to warrant the additional run time. So far the main value of the Indirect Method, to the author, has been to act as an independent check on the Direct Method.

- The Indirect Method has the technical advantage of being more widely applicable and uses very little theory. The Direct Method requires knowledge of the copula generating the data and becomes more difficult to work with if a risk measure other than *CTE* is used.

- The Indirect Method has the technical disadvantage of requiring more run time, and it may run into issues as the size of the optimization problem increases.