Revisiting Copula Dependency Modeling:
A Case for Conservatism[1]

Hulusi Inanoglu, OCC
Scott Ulman, FDIC
April, 2009

Part I. Introduction

The Advanced Measurement Approach (AMA) of the new U.S. Rule implements Basel II initiatives. The Rule requires large banks to calculate their total operational risk exposure (ORE) by aggregating individual loss exposures calculated for *units of measure*. Each separate unit of measure must represent an appropriate level (e.g. business unit or operational loss event type) at which the bank's quantification system generates a separate probability distribution for potential operational losses. Under the Rule, units of measure may not combine business activities or operational loss events with *demonstrably different* risk profiles within the same loss distribution.

To calculate total ORE, qualifying U.S. banks must aggregate individual operational-risk loss exposures across units of measure. Most bankers hope that increasing the number of units of measure will inevitably increase "diversification" and dramatically reduce the total ORE via a portfolio-type effect as in market-risk. Under the Rule, ORE is measured in terms of a statistical measure known as operational-value-at-risk, OpVaR(L;α) which is a quantile function intuitively indicating the smallest operational-risk loss L at which the cumulative distribution function is greater or equal to the α-percentile.[2] Alternatively, ORE is the smallest loss such that the survival function is less or equal to the (1-α)-percentile.

Diversification clearly depends upon multivariate dependence which is rather vaguely defined in the Rule as

> a *measure of the association* among operational losses across and within units of measure[3].

Furthermore the Rule states that the agencies expect that a bank's assumptions regarding dependence ***will be conservative given the uncertainties surrounding dependence modeling for operational risk***. If a bank does not satisfy the requirements surrounding dependence, the bank must sum operational risk exposure estimates across units of measure to calculate its total operational risk exposure.

Finally, the Rule states that banks

> may use *internal estimates of dependence* among operational losses within and across business lines and operational loss events if the bank can demonstrate to the satisfaction of its primary federal regulator that its process for estimating dependence is sound, robust to a variety of scenarios, and implemented with integrity, and allows for the uncertainty surrounding the estimates. If the bank has not made such a demonstration, it must sum operational risk exposure estimates across units of measure to calculate its total operational risk exposure.[4]

The Rule's requirement (in the absence of a sound, robust dependence model) to sum individual $OpVaR_\alpha(L_i)$ across units of measure probably indicated regulators' initial **hope** that the sum on the right-hand side might provide an *upper bound* to aggregate annual losses at some high (α=0.999) percentile as shown on the RHS of equation (1a):

---

[1] The opinions expressed in this paper are those of the authors and do not necessarily reflect the judgment of FDIC or OCC. Email the authors at SUlman@fdic.gov or hulusi.inanoglu@occ.treas.gov. This is a preliminary version of the paper since it does not yet include tables of results analyzing LDCE2008 data for multiple banks which did not become available to the authors until mid-March 2009. All empirical results in this paper were produced using R-language software written by the authors or taken from the R-language version of QRMLib which supports the risk-management textbook by McNeil, Frey, Embrechts (2005).

[2] Technically, this is written as $OpVar_\alpha(L_i) = \inf\{L_i : F(L_i) \geq \alpha\}$ where α is a high percentile value like 0.999 (99.9%) and $F(L_i)$ is the cumulative distribution function of the losses. . Alternatively, $OpVar_\alpha(L_i) = \inf\{L_i : 1 - F(L_i) \leq 1 - \alpha\}$.

[3] A unit of measure means "the level (for example, organizational unit or operational loss event type) at which the [bank]'s operational risk quantification system generates a separate distribution of potential operational losses." Under the U.S. Rule, a unit of measure can "not combine business activities or operational loss events with demonstrably different risk profiles within the same loss distribution". Using a model with only a few units of measure increases the need to insure potential measurement of any loss associations within each single unit of measure.

[4] Taken from the Final Rule, section 22 (h)(3)(D)

(1a) $\lim_{\alpha \uparrow 1} OpVaR_\alpha(\sum_{i=1}^d L_i) \le \lim_{\alpha \uparrow 1} \sum_{i=1}^d OpVaR_\alpha(L_i)$

Verbally, equation (1a) may be restated as

(1b) $totalOpVar \le subadditivity\ upper\ bound$

When considering only two units of measure (d=2) in (1a), the upper bound on the RHS is known as the case with perfect positive dependence or comonotonicity since the two risks then have maximal linear correlation. The general case where the sum of OpVaRs across individual units of measure accurately determines the upper bound on aggregate losses is said to provide a **subadditive** measure.

Subadditivity for portfolio VaR has been extensively investigated in market risk and is frequently interpreted as implying that "diversification effects" almost invariably reduce total VaR below the subadditivity-bound for aggregated returns or losses.[5] In fact, a measure of **diversification** recently introduced by Embrechts, Lambrigger, and Wuthrich (2008) measures the difference between the left- and right-hand sides of equation (1):

(2a) $D_\alpha(L_1 + ... + L_d) = OpVaR_\alpha(\sum_{i=1}^d L_i) - \sum_{i=1}^d OpVaR_\alpha(L_i) = totalOpVaR - subadditivity\ upper\ bound$.

A negative value for D_α indicates the amount by which diversification reduces the total loss since the ORE across the summed losses is less than the subadditivity upper bound in (1)).

An alternative measure shown later in tables of this paper is the *diversification ratio* or *capital reduction ratio*

$$\begin{aligned} DivRatio_\alpha &= (\textstyle\sum_{i=1}^d OpVaR_\alpha(L_i) - OpVaR_\alpha(\sum_{i=1}^d L_i)) / OpVaR_\alpha(\sum_{i=1}^d L_i) \\ &= (\textstyle\sum_{i=1}^d OpVaR_\alpha(L_i) / OpVaR_\alpha(\sum_{i=1}^d L_i)) - 1 \qquad (2b) \\ &= (subadditivity\ upper\ bound / totalOpVaR) - 1 \end{aligned}$$

which measures the proportion by which the ratio of the subadditivity upper bound to the total OpVaR exceeds one. This effectively represents the proportion of capital reduction provided by "diversification" effects. In cases where the *totalOpVaR* exceeds the subadditivity upper bound, the diversification ratio (capital reduction) will be negative. This case is known as superadditivity (where diversification fails).

Consequently, the study of diversification in risk management is integrally related to determining conditions where OpVaR fails to be subadditive. Furthermore, aggregation of losses, multivariate dependency, diversification, and subadditivity are interlinked in some way and, hopefully, various dependency measures can be derived or separated to explain diversification effects.

Probably relying on the notion of subadditivity as an “upper risk bound”, several banks during the NPR period of the U.S. Basel II Rule commented that the Rule’s requirement (in the absence of a validated model) to sum OpVaR across units of measure (effectively using the RHS of equation (1)) is *excessively conservative* since it assumes *perfect positive dependence*. The complaints probably reflected bankers’ experiences with market risk analysis where log-returns on stocks and exchange rates are usually viewed as subadditive due to portfolio diversification benefits.

Traditional dependency modeling in market-risk frequently utilizes symmetric elliptical[6] return distributions with two-sided support in which case (Pearson's) linear correlation statistic provides the sole required dependence measure. McNeil, Frey, Embrechts (2005) have proved that portfolio

[5] Although the elliptical model does well in a large majority of finance models, the literature does describe multiple situations where subadditivity fails. These include situations like portfolios with very skewed loss distributions (possibly containing options or defaultable bonds), assets with infinite-mean marginal distributions, or loss distributions created from highly asymmetric copulas.

[6] Elliptical distributions are frequently used for log returns on stocks and exchange-rate returns, especially when observations are weekly or more frequent. They provide significant improvement in fit over Gaussian returns. The elliptical family includes multivariate normal and multivariate normal-variance mixtures including the multivariate-t and the symmetric multivariate NIG distributions. The normal-variance mixtures thicken the tails and allow joint tail dependency if the mixing variable has a distribution with a power tail. Normal mean-variance mixtures are not elliptical since the mean-transform skews the distribution, creating possible asymmetries in the tails. For a thorough description of elliptical models, see McNeil, Frey,

> "VaR is subadditive when all portfolios can be represented as linear combinations of the same set of underlying elliptically distributed risk factors....an elliptical model may be a reasonable approximation for various kinds of risk-factor data such as stock or exchange-rate returns."

In response to bankers' NPR comments requesting a greater diversification benefit than provided by summing OpVaR across units of measure, the Final Rule stated

> "the agencies continue to believe that this treatment of operational risk exposure estimates across units of measure is prudent until the **relationships among operational losses are better understood**…. The agencies expect that a bank's approach for developing explicit and objective dependence determinations will **improve over time**."

The preceding statements and responses to comments in the Rule about ***conservative modeling***, ***better understanding*** in the future, and ***improvement over time*** imply U.S. regulators believe that *dependency modeling in operational risk* is much more complicated and less well-developed than the successful and popular paradigms previously developed for market risk (and credit risk) and is not yet thoroughly understood. Clearly scarcity of operational-risk loss-data due to the annual loss horizon is one factor which critically distinguishes operational-risk modeling from the market-risk world. However, other major drivers distinguishing the operational-risk model from the market-risk model include vast differences between the *support* and the *tail-behavior* of operational-risk losses when compared to market-risk returns.

Well before finalization of the Rule for operational risk, academic and regulatory investigators had published evidence using EVT (Extreme Value Theory) to estimate capital required to protect against losses associated with heavy-tailed distributions with one-sided support $(X \in (0,\infty))$ assumed to characterize operational-risk losses. Several investigators[7] reported preliminary estimations for OpVaR greatly exceeding the subadditivity upper bound. Such "superadditivity" violations are usually consistent with infinite-mean models associated with "extremely heavy-tailed models[8]". Furthermore, the risk-management textbook of McNeil, Frey, Embrechts (2005) had already proved that the comonotonic subadditivity bound in (1a) does **not** always hold for VaR and hence the maximum VaR for two assets can correspond to a smaller linear correlation than the attainable maximum of 1 under perfect correlation.[9] Hence the case of perfect positive dependence **cannot** accurately be described as excessively conservative.

Embrechts (2005) text, Sections 3.3.2 and 3.3.3. Technically, an elliptical rv is a multivariate affine transformation of a spherically distributed variable Y such that $X \overset{d}{=} \mu + AY$. A spherical distribution is a multivariate distribution with uncorrelated components and identical, symmetric margins, e.g. $Z \sim N_d(0,I_d)$. An example of an elliptical normal-variance mixture is the student-t where $Y = \sqrt{W}Z$ creates the mixture $X \overset{d}{=} \mu + \sqrt{W}AZ$. The student-t mixing variable W is distributed as an inverse-gamma which has a heavy tail which in turn implies joint tail dependence.

[7] See especially studies by Moscadelli (2003), de Fontnouvelle et al (2004), Dutta and Perry (2006).

[8] In this document we divide loss distributional tails into five intuitive types: "extremely-heavy" implying an infinite-mean model; "heavy-tailed" implying an infinite-variance model; "medium-tailed" implying a model with all moments finite and tail heavier than exponential; "light-tailed" models whose tails decrease at an exponential rate or faster; "right-truncated" indicating a model with a finite right endpoint.

[9] The McNeil, Frey, Embrecht's (2005) textbook calls this Fallacy 3 regarding linear correlation. The other two fallacies included 1) marginals and pairwise linear correlations determine the joint distribution in all cases; 2) for two different marginals and a selected linear correlation coefficient $\rho \in [-1, 1]$, it is always possible to construct a joint distribution. Linear correlations are invariant only under linear transformations. Under nonlinear transformation, the ρ value may be extremely constrained to an interval like [-0.5, 0.5]. For example, consider two lognormal variables with distributions $\ln X_1 \approx N(0,1)$ *and* $\ln X_2 \approx N(0,\sigma^2)$. Then ρ = 1 is possible only if σ = 1. Otherwise for σ > 1, the maximum linear correlation value may become significantly less than 1. For example, the **maximum** linear correlation value is ρ = 0.5 if σ = 2.4. As σ rises, feasible values of ρ will fall (even approach 0) even though the two lognormal variables are not independent.

Despite the knowledge that VaR may be superadditive, regulators adhered to the subadditivity upper bound to set the minimum capital requirement (MCR) representing ORE. Obviously regulators expect (hope) that violations inducing superadditivity discovered in early models were caused by inappropriate model assumptions or measurement error (like mixing data with different risk characteristics) which can be overcome by more careful modeling.

Unfortunately, the Final Rule did not elucidate which factors in modeling operational-risk dependencies regulators believe to be "misunderstood". Part II of this paper will attempt to summarize several areas of operational-risk dependency modeling which seem to be realtively unresolved and hence require ongoing conservative treatment as suggested by regulators' comments in the Rule. Conservatism contrasts with many banks' desires to assume that operational risk capital can be immediately and significantly reduced (by as much as 50%) through diversification by assuming that loss distributions have "thin-tails". In particular Part II will provide compelling and intuitive graphical evidence that the one-sided support provided by operational-risk loss distributions precludes much of the diversification benefit provided by two-sided returns in the capital markets. Furthermore, it will show that interactions between characteristics of both marginal distributions and the copula (joint distribution) create unanticipated impacts on diversification. Hence analysts should not expect operational-risk losses to attain large diversification benefits which significantly reduce the required capital level below the subadditivity upper bound. Furthermore, any scenario analysis implemented for operational-risk dependency cannot be based on logic derived from market- or credit-risk where underlying data has two-sided support

Part III utilizes simulation via a distributional copula model to calculate the diversification ratio in equation (2b) and other pertinent measures like VaR and expected shortfall (ES). A hybrid version of the two-step inference-for-marginals (IFM) method will be utilized to separately estimate parameters for the marginals and the distributional copula. Historical operational-risk losses will be aggregated across event-types (ET) proposed by Basel II. We believe that aggregation across business lines is generally not conservative since it is highly susceptible to combining data with different risk profiles into a single distribution in violation of the U.S. Rule. Furthermore, historical bank data will be aggregated on a weekly basis to avoid data synchronicity problems since data in some event loss types is relatively sparse. The marginals in this study will be fitted by MLE as *mixtures* of a "medium-tailed" lognormal distribution in the body and a "heavy-tailed" GPD in the tail. Our study will show that some methods of initializing estimators and calculating an empirical threshold in the internal MLE process can cause significant underestimation of the vital shape parameter in the GPD tail. The study also develops equations for the density functions of the Gumbel copula in up to five dimensions The density equations are needed to estimate parameters via loglikelihood methods for the conservative Gumbel copula with its positive tail dependence which may reduce diversification. Once the estimations of marginal-distribution and copula-parameters are completed, joint annual losses over a 100,000 year horizon can be simulated from the distributional copula in an approach analogous to the Loss Distribution Approach (LDA). Resulting VaR, Expected Shortfalls (ES), diversification ratios (see equation (2b)) and other statistics at high quantiles can then be inferred from the simulated data. We show meta-distributional results estimated from Gaussian, t, and Gumbel copulas with the potentially heavy-tailed marginal mixture distributions.

Part II. Dependency Relationships Under Scrutiny

A. General Copula and Marginal Characteristics

The Final Rule calls for conservative treatment until **relationships among operational losses are better understood.** In this section, we discuss dependency issues which may need further study and clarification. We also present clear graphic evidence that the absence of negative tail dependency in operational-risk losses (which necessarily have only one-sided support) means that diversification effects cannot be nearly as strong as in market risk. This implies that conservative scenario analyses cannot mirror diversification effects in market or credit risk.

Recent operational risk literature[10] is replete with studies showing that VaR (and OpVaR) can be **superadditive.**

[10] "Multivariate Extremes and the Aggregation of Dependent Risks: Examples and Counter-Examples", *Extremes* 2008, Embrechts, Lambrigger, and Wuthrich references much of the relevant literature.

Specifically, superadditivity prevails whenever the VaR measure for the aggregate (across units of measure) annual loss exceeds the subadditivity upper bound as shown in equation (3).

(3) superadditivity $\Rightarrow \lim_{\alpha \uparrow 1} OpVaR_\alpha(\sum_{i=1}^{d} L_i) > \lim_{\alpha \uparrow 1} \sum_{i=1}^{d} OpVaR_\alpha(L_i)$

Superadditivity implies a *lack of diversification* (in fact, negative diversification) and brands OpVaR as an incoherent risk measure according to Azner.

Since U.S. regulators chose to support OpVaR even when they knew about violations of subadditivity, it is clear they must be very interested in determining what causes the superadditivity violations and finding adjustments to eliminate the underlying causes. Recent research suggests *three* important factors related to the marginal distributions and *two* important factors related to copulas may **interact** to reduce diversification results; they must be examined carefully to see if their interaction might lead to superadditivity.

Marginal Distribution Factors

- *Tail behavior (variation) of the marginals*.
  Researchers initially suspected that infinite-mean models (where the tail is so heavy that no constants for convergence can be found) might be the primary cause of superadditivity. Infinite-mean models were routinely empirically identified in operational-risk banking research and are known as "extremely heavy tailed" models. Some researchers suggested that fittings of infinite means to data were caused by data contamination (mixing data from distributions with different risk-characteristics) rather than correctly categorizing a loss distribution. A possible remedy is to rationally filter outliers. However, considerably more work needs to be done in this area.
- *Symmetry or asymmetry of the underlying marginals*
  The symmetry factor interacts with the "extremely heavy tailed" factor, the two-sided support factor, and the tail-dependence factor for some copulas. Embrechts, Lambrigger, and Wuthrich (2008) show that subadditivity prevails for elliptical (symmetric) marginals whose tail variation implies an infinite-mean model as long as the data support for the marginals is two-sided. Hence, contrary to intuition, infinite-mean models are not necessarily superadditive. They write:
  > "For the infinite mean multivariate *t*-distribution, subadditivity of VaR holds due to the dependence properties in the upper left and lower right corner[11]. High values of one risk are compensated by low values of the other risk, turning VaR into a coherent risk measure for such infinite mean models."
- *One-sided or two-sided support for the marginals*[12].
  With one-sided support (as in operational risk where losses cannot be negative) elliptical models with infinite means become superadditive even though they are subadditive with two-sided marginals. One-sided support precludes the negative tail dependencies (where large losses are offset by large positive gains) described in the previous quotation.

Copula Dependence Factors

- Dependence structure cannot be described solely by correlations (rank or linear) for most copulas.
- Tail Dependency
  - Positive Tail Dependency
  - Negative Tail Dependency (precluded in distributions with one-sided support)

Positive and negative tail dependence both depend on the copula (joint density) and not on the marginals. However, some copulas (like the Gaussian) have tail dependence coefficients asymptotically equal to 0. A more detailed description of tail dependency and its impact on diversification is provided in the next section.

B. Impact of Tail Dependency on Diversification

[11] Upper left refers to upper negative tail dependency; lower right refers to lower negative tail dependency. These concepts will become clear in the graphs which follow.

[12] Data are said to have one-sided support if their values lie on the interval $(0, \infty)$ like operational-risk losses or two-sided support if their values lie on the interval $(-\infty, \infty)$ like stock returns.

It is vitally important to understand the concept of tail dependency now that regulatory capital has been restructured to protect against "one in a thousand year" cataclysms for operational risk. Clearly, tail dependence is also critical in explaining ordinary portfolio diversification for market risk on daily or more frequent samples. Tail dependence in operational risk measures the tendency of large outcomes (tail events) to occur "simultaneously, i.e. within the same year" in two (or more) loss ETs. Tail dependence measures do not depend on the underlying marginals and hence can be written in terms of the joint distribution's underlying distributional copula. See Appendix 1 for technical definitions of tail dependence. We will first provide an intuitive description of tail dependency and then provide graphics to give an intuitive explanation of its effects on diversification.

In the following description, a "variable" refers to an annual operational-risk loss within a distinct unit of measure and "simultaneously" should be interpreted to mean "in the same time period as the risk-analysis horizon".

Positive tail dependency λ^+ :
Positive tail dependency Is the probability that within the loss year, both variables will exhibit **similar behavior** in the tails. There are two separate types of positive tail dependency:

- Positive upper tail dependence λ_u^+ : the probability that both random variables will produce similar large losses in the upper tail (largest quantiles: upper-right region) within the year. This is extremely important in operational risk since it reflects the opposite of diversification.
- Positive lower tail dependence λ_l^+ : the probability that both random variables will produce similar small losses in the lower tail (smallest quantiles: lower left region) within the year. It is basically irrelevant for diversification in random variables with single-sided infinite support $X \in \Re = (0,\infty)$ like operational-risk losses since it describes only joint outcomes which are both close to zero.

Negative tail dependency λ^- :
Is the probability that within the loss year both random variables will simultaneously exhibit **opposite behavior** in the tails. This concept is a major driver of diversification in market risk since it implies that very large losses in one tail can be offset by very large gains in the opposite tail. The concept is relatively meaningless in operational risk where single-sided infinite support $L \in \Re = (0,\infty)$ for losses means that a large loss in one unit of measure is "offset" by a loss close to zero in another unit of measure since large negative losses cannot occur.

- Negative upper tail dependence λ_u^- : the probability that one random variable will produce an outcome in the upper tail (highest quantile: upper left region) and the other random variable will produce a simultaneous outcome in the lower tail (smallest quantile: upper left region) within the loss year
- Negative lower tail dependence λ_l^- : the probability that one random variable will produce an outcome in the lower tail (smallest quantile: lower right region) and the second random variable will produce a simultaneous outcome in the highest tail (largest quantile: lower right region) within the loss year

A copula may have

- both upper and lower tail dependence (t-copula, Generalized Clayton copula)
- upper tail dependence only (Gumbel copula)
- lower tail dependence only (Clayton copula)
- no tail dependence (Gaussian copula, Frank copula)

Effective negative tail dependency is reliant upon *two-sided* data support which is a typical description for variables which can take on both positive and negative values. Infinite two-sided support describes the variables' ranges as being $X \in (-\infty,\infty)$. For example, stock-market log-returns exhibit two-sided support. Contrarily, operational-risk losses have only one-sided support (their range is $X \in (0,\infty)$) since "negative" or "zero" operational-risk losses are undefined.

Major diversification benefits are provided by copulas exhibiting **negative** tail dependencies when the pair of X variables has two-sided support: large positive values from one variable can offset large negative values from the other variable, a concept often recognized as portfolio diversification.

Figure 1 presents a pair of plots representing simulations from two bivariate t-copulas with identical underlying t-marginals. One bivariate copula has a positive correlation coefficient; the other copula has a negative correlation

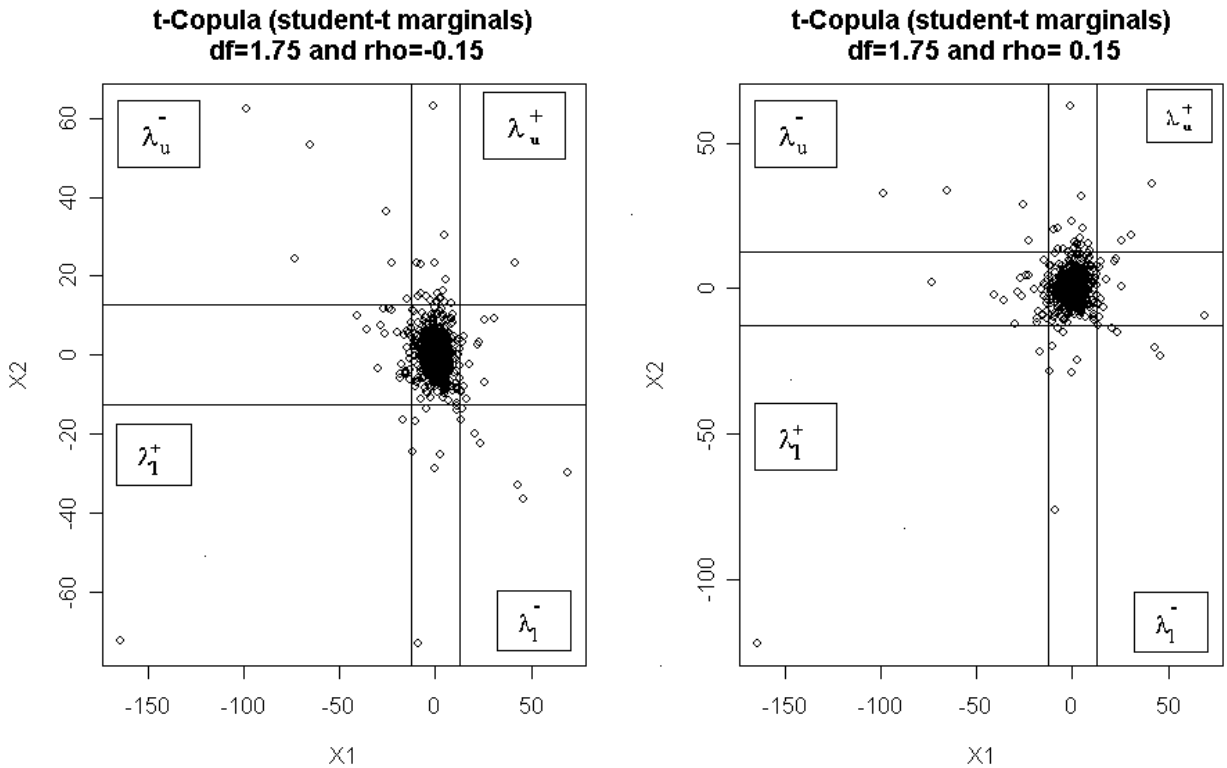

Figure 1. With two-sided support, simultaneous large negative and positive values are possible) for both X_1 and X_2. The left-side graph shows the straight t-copula with small negative correlation versus the right-side graph with small positive correlation. The horizontal and vertical lines represent .005 and .995 percentiles. Note the nine separate regions where losses can occur with two-sided support for the marginals.

In both graphs, the segments marked with **negative** tail-dependency (upper (λ_u^-) or lower (λ_l^-)) clearly show that large positive values from X_1 occur simultaneously with large negative values from X_2 and vice versa when X has t-marginals. The simultaneous large positive values offset the large negative values, a phenomenon frequently driving diversification.

Note the t-copula also contains a few cases of positive upper (λ_u^+) and lower (λ_l^+) tail dependency in the uppermost right and lowermost left regions which create the opposite effect of diversification.

coefficient. Changing the sign of the correlation coefficient rotates the bulk of points in the center of the plots. It may also have reduced the number of outcomes with positive upper tail dependency λ_u^+ but did not totally eliminate such simultaneous extreme observations.

Since t-distributions have two-sided support, the addition of q=.995 and 1-q = 0.005 percentile lines for both variables displays **nine** feasible regions for outcomes in the copula plot. Negative tail dependency λ^- is associated with both the upper-left and lower-right of the nine regions. Hence negative tail dependency has both upper and lower components. It is the basic explanatory power behind subadditivity prevailing even if the tail variation implies infinite-mean models as shown by Embrechts, Lambrigger, and Wuthrich (2008):

> "For the infinite mean multivariate *t*-distribution, subadditivity of VaR holds due to the dependence properties in the upper left [λ_u^-] and lower right corner [λ_l^-]. High values of one risk are compensated by low values of the other risk, turning VaR into a coherent risk measure for such infinite mean models."

In particular, Figure 1 clearly shows that the standard t-copula generates numerous data pairs with large negative outcomes being offset by simultaneous large positive outcomes in both upper left and lower right regions of the

graph. These points play a crucial role in diversification since the variables have doubly infinite support. The graph shows negative tail dependency only between two assets. When hundreds of assets exhibiting similar negative tail dependencies are combined in a portfolio, the diversification impact may be greatly magnified.

For infinite-mean models with two-sided support, negative tail dependency can be so strong that positive extreme values in one coordinate are sufficiently compensated by negative extreme values in the other coordinate to maintain subadditivity. In particular Embrechts, Lambrigger and Wuthrich (2008) have proved elliptical distributions maintain asymptotic subadditivity for infinite-mean models as long as they exhibit two-sided support. The absence of negative tail dependency induced by one-sided support of the marginals intuitively precludes large negative values from offsetting (diversifying) large positive losses.

Despite the fact that two-sided support is *necessary* for negative tail dependencies with their remarkable diversification benefits, Figure 2 shows that such support is *not sufficient*. The copula-type also plays a vital role.

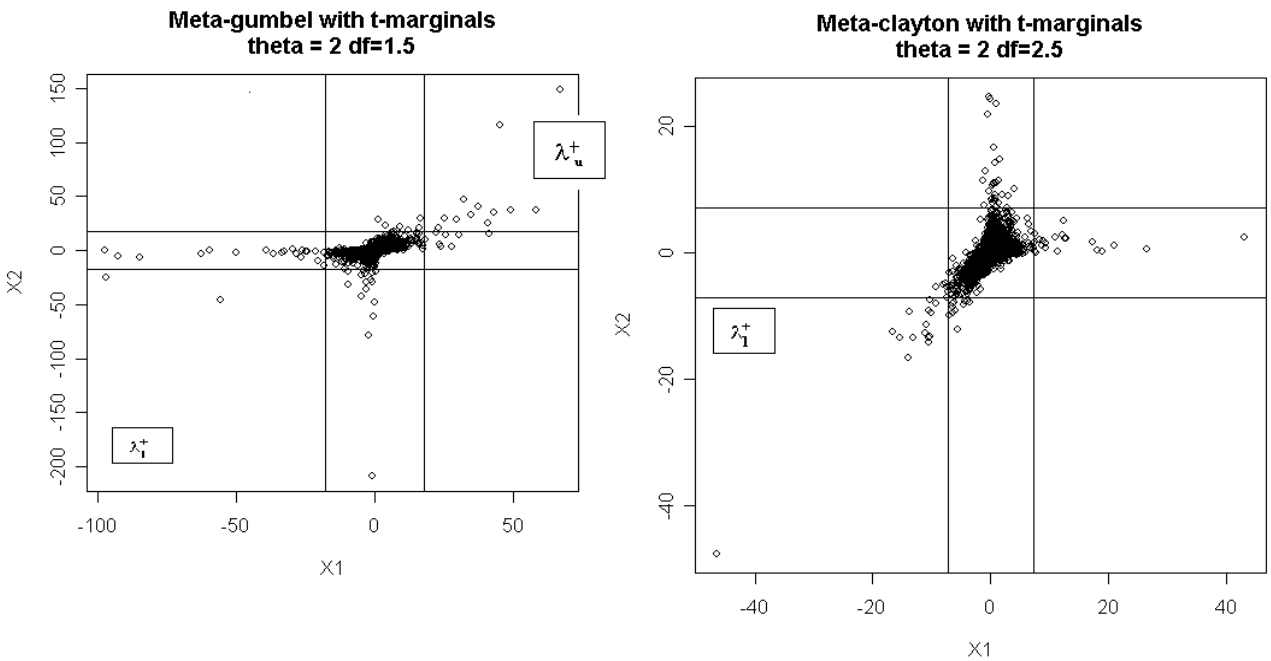

Figure 2. The graphs show simulated meta-gumbel and meta-Clayton copulas with parameter θ=2 and underlying student-t-marginals with 1.5 degrees of freedom. Neither the gumbel nor Clayton copula has any **negative** tail-dependency which is verified by the absence of any simulated points in the upper left or lower right regions in both graphs. Hence large positive values from X_1 never occur simultaneously with large negative values from X_2 and vice versa which greatly reduces any "diversification effects".

In general, the gumbel copula is said to exhibit only positive upper tail dependency λ_u^+ implying situations where large positive events occur simultaneously. Positive tail dependencies are clearly the opposite of diversification. Similarly, the Clayton copula is said to exhibit only positive lower dependence where large negative outcomes occur simultaneously for both variables but large positive outcomes do not occur simultaneously.

Despite having two-sided support for the marginals, both copulas totally lack the negative tail dependency which is intuitively the driver of diversification in portfolio theory.

The horizontal and vertical lines represent .005 and .995 percentiles

Certain Archimedean copulas totally preclude negative tail dependencies even when the marginals have two-sided support. The left-hand graph in Figure 2 shows values simulated from a Gumbel copula with student-t marginals (a

so-called meta-Gumbel distribution with t-marginals). The right-hand graph shows values simulated from a Clayton copula with student-t marginals (a so-called meta-Clayton distribution with t-marginals). Neither graph exhibits any negative tail dependencies (in the upper left or lower right regions where large negative values offset simultaneous large positive values) despite the marginals having two-sided support. Additionally, the Clayton copula exhibits no positive upper tail dependence but does show positive lower tail dependence indicating a propensity for large simultaneous negative values. Contrarily, the Gumbel copula exhibits positive upper tail dependence, indicating a propensity for simultaneous positive values. This is beneficial if the "values" are stock returns but it is dangerous if the "values" are operational-risk losses.

The existence and importance of tail dependencies change dramatically when considering operational-risk losses which exhibit one-sided support (are defined as $X \in (0, \infty)$). The absence of two-sided support totally precludes any meaningful negative tail dependencies and hence implies a serious decline in meaningful "diversification" benefits for large losses. Furthermore, presence of positive tail dependency in the underlying copula would further reduce any potential diversification.

Figure 3 shows losses simulated from a t-copula with lognormal marginals (a meta-t distribution with lognormal

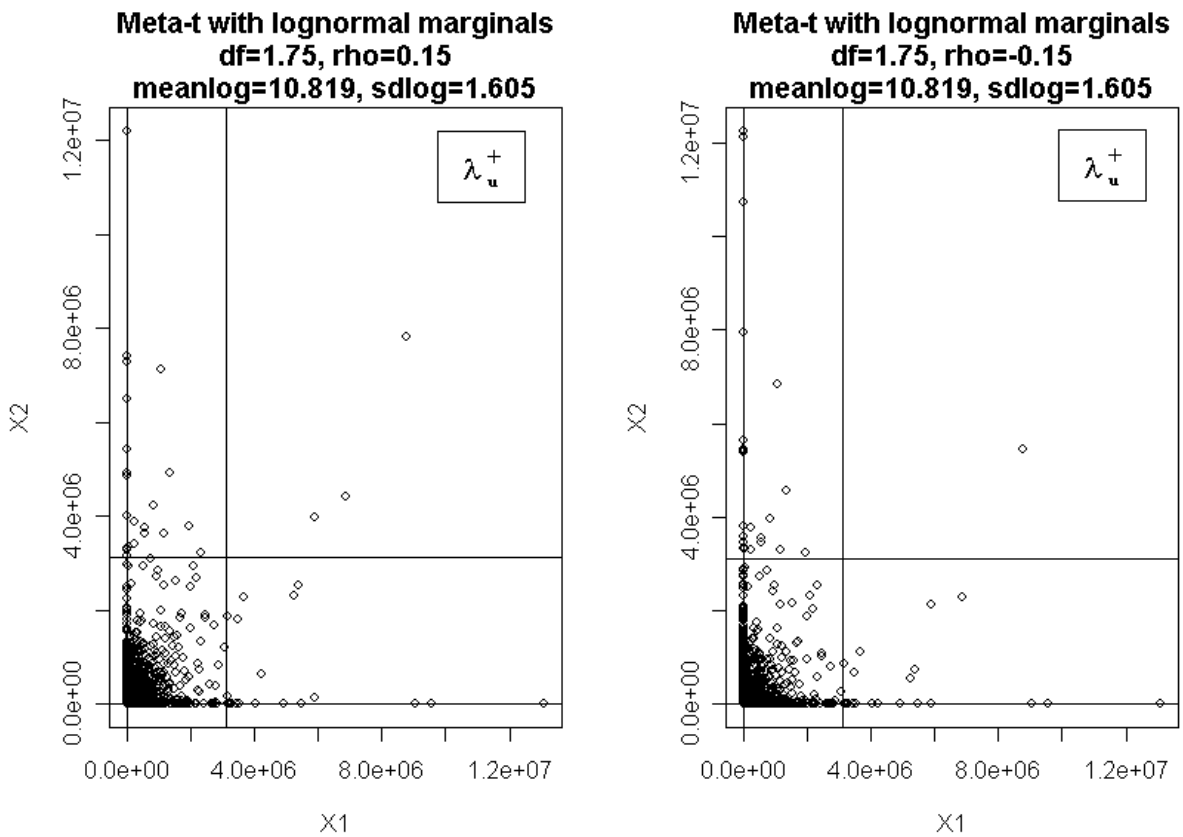

Figure 3. Simulations of bivariate losses from a t-copula with identical lognormal marginals. The horizontal and vertical lines represent .005 and .995 percentiles.

Note the total absence of negative tail dependence which is a primary driver of diversification in market risk since it allows large negative and positive outcomes to offset one another. Such diversification is impossible in loss distributions with one-sided support. In fact, note that only ***four*** of the ***nine*** **regions** now contain potential outcomes due to the one-sided support of the lognormal marginals.

Note that the negative correlation in the second figure may have reduced the number of loss pairs with upper tail dependency but did not eliminate them entirely.

marginals). Note that the best outcome for λ_u^- is now a simultaneous loss of zero coupled with another positive loss since operational-risk losses cannot be negative. The absence of negative support seriously curtails the potential for diversification when extreme losses occur. In fact, Figures 3 and 4 demonstrate that one-sided support for the

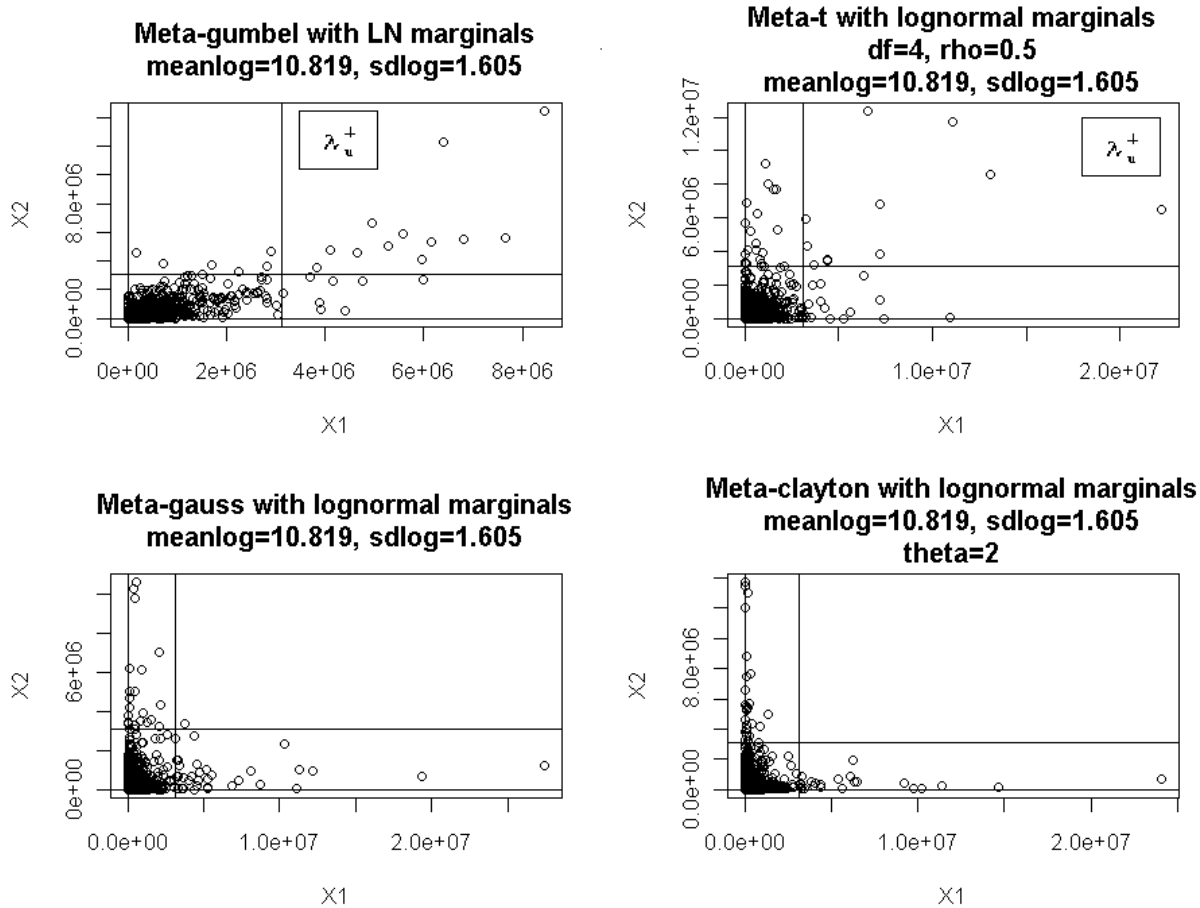

Figure 4. Simulated losses from four different bivariate copulas each with identical lognormal marginals (characterized by one-sided support).
Note that only four of the nine regions in the graph contain simulated values.
The horizontal and vertical lines represent .005 and .995 percentiles

marginals effectively reduces the number of regions potentially containing losses from the nine regions shown in Figures 1 and 2 to only four regions in Figures 3 and 4. The five regions on the bottom and left-hand side of the plot can have no observations when the margins have one-sided support. The omitted upper-left and bottom-right regions generally contain data pairs associated with negative dependency.

Figure 4 shows graphs of four different meta-distributions with marginals all exhibiting one-sided support. Each meta-distribution of course lacks any meaningful *negative* tail dependency. The importance of the simulated distributions in Figure 4 is that some copulas also preclude *positive* tail dependence. Neither the Gaussian nor Clayton copula exhibits any large simultaneous tail events. Hence usage of these copulas guarantees a lower level of operational-risk losses from simultaneous tail events and may not be a conservative treatment. Thus it becomes very important to validate a correct copula choice.

To summarize the results of this section, it is important to emphasize that in terms of copula characteristics, rank correlation cannot entirely describe dependency in a risk model unless the underlying distributions exhibit elliptical symmetry. Even in the absence of significant rank correlation, tail dependency in the copula may result in either increased or decreased diversification benefits since it may be either positive or negative, just like correlation. Negative tail dependency provides major explanatory power for “diversification” benefits while positive tail dependency reflects simultaneous large losses or gains which reduce "diversification". The fact that operational-

risk losses are restricted to positive one-sided support for the marginals can be considered as an implicit restriction on the dependency structure which *precludes all negative tail dependency*. The absence of negative tail dependency due to one-sided support prevents large negative losses from offsetting (diversifying) large positive losses. For example, a standard t-copula with t-marginals allows negative tail dependency with a potential improvement in diversification. However, if the propose marginals are switched to lognormal distributions with one-sided support (creating a meta-t copula), the possibility of "negative" losses is immediately eliminated along with any major "diversification" potential.

Part III. Measuring the Diversification Effect

A. Choosing a methodology to measure the diversification ratio

Over the past decade, researchers have proposed several methodologies to model dependencies in loss distributions. Models based on *simulations of copulas* have become predominant since the copula allows separation of dependency factors from the underlying loss distributions when trying to extract diversification effects. There are currently two separate copula approaches to dependency modeling.[13]

1) Model dependence by using **distributional copulas** where the aggregate loss distribution for each unit of measure is assumed to be a compound Poisson process. Distributional copulas were first popularized by Sklar in 1959. A *distributional* copula[14] is a joint distribution function $C(u_1,...u_d)$. with standard uniform marginals. Hence the distributional copula maps the unit hypercube into the unit interval as $C(\mathbf{u}):[0,1]^d \rightarrow [0,1]$. The distributional copula is amenable to translation using both quantile transforms and probability transforms. Sklar (1959) proved that for uniform marginal distributions representing probability transforms ($u_1 = F_1(x_1),..., \quad u_d = F_d(x_d)$) of continuous marginals that the resulting joint distribution function is unique:

$$C(u_1,...u_d) = C(F_1(x_1),..., F_d(x_d)) = F(x_1,..., x_d).$$

Hence the copula function separates modeling the dependency structure from the marginals. Furthermore, changing marginal distributions implies changing the joint distribution even if the copula function remains unchanged. Thus the converse of Sklar's Theorem implies that *arbitrary marginals* may be applied with a copula specification to create alternative joint distributions with similar dependency traits.

It is vital to understand that marginal densities associated with distributional copula models do not need to come from the same "family" as the joint distribution. The joint distribution created from a t-copula using univariate student-t marginals is indeed a multivariate student-t distribution. However, the joint distribution created from a t-copula combined with lognormal marginals is known as a ***meta-t** distribution* (or a meta-t copula) with lognormal marginals. It has dependency characteristics of the multivariate-t but marginal characteristics of the lognormal distribution. In general, a "meta-x-distribution" is a joint distribution created from an "x"-copula model using arbitrary margins.

The diversification effect can then be measured by a procedure which estimates the marginal distributions, estimates the copula parameters, then simulates annual losses across units of measure (loss event-types) using parameters from the prior fittings over a very large time period. The VaRs for the simulated annual losses may then be obtained from the empirical distribution function of the simulated losses for each unit of measure and equation (2b) can be applied to calculate a diversification benefit.

---

[13] Three other methods do not use copulas. They include i) modeling frequency dependency between loss-occurrence times using linear (Pearson's) correlation; ii) modeling severity dependency for losses occurring simultaneously; iii) modeling coincident loss events using a "common-shock" model. Bock and Kluppelberg (2007) show that two factors severely limit the importance of modeling frequency correlations in compound Poisson models: 1) Since the loss-frequency dependency affects only the number of expected losses but not the size of the losses, it has a very limited impact on the size of total aggregate loss in units of measure when some losses are very large. 2) If two units of measure have mean frequencies with highly different levels, then the resulting frequency correlation value is limited to very low values.

[14] Most risk-management and statistics textbooks currently refer to "distributional" copulas as copulas. However, in 2007-8, Bocker and Kluppelberg have developed the Levy-copula (and Pareto-Levy copula) as methods for extracting dependency information. Hence we have chosen to use the distributional adjective to clarify the type of copula under discussion.

2) A second alternative would be to model dependence via **Levy** or **Pareto-Levy copulas (**which are not distributional copulas since they depend upon tail integrals). A Levy copula is associated with multivariate compound Poisson distributions (finite jumps in a time interval) while the Pareto-Levy copula is associated with a process having infinite variation (an infinite number of jumps in the time interval). In 2007-8, Bocker and Kluppelberg developed Levy- and Pareto-Levy copulas as methods for extracting dependency information for multivariate loss distributions. The Levy copula separates marginal tail integrals and dependencies in jump measures from the joint tail integral somewhat analogously to distributional copulas via the following definition: $\overline{\Pi}(x_1,...x_d) = C^{Levy}(\overline{\Pi}_1(x_1),,,\overline{\Pi}_d(x_d))$ where $\overline{\Pi}(x_1,...x_d)$ represents the joint tail integral and $\overline{\Pi}_i(x_i)$ represents the ith marginal tail integral. Hence a Levy copula maps $(0,\infty]^d \rightarrow [0,\infty]$ rather than mapping from the unit hypercube into the unit interval and hence is not a distributional copula. See Appendix 3 for further details.

This paper chooses the **distributional** copula dependency models approach where distributional copulas measure dependency across severity distributions only (ignoring frequency dependencies). We seek to measure and compare the "diversification effects" for several banks using LDCE 2008 data, i.e. their reduction in ORE implied by dependency factors from assumed distributional copula models.

B. Resolving Data Synchronicity and Dependency Problem

A major problem with measuring dependency using a distributional copula approach is the total scarcity of annual operational-risk loss data within individual units of measure for a single bank. Hence data must be aggregated on a shorter time horizon (monthly, weekly, or daily). The problem is further compounded because the marginal distributions for operational-risk losses are recognized to be heavy-tailed, frequently requiring application of Extreme Value Theory (EVT) for parametric estimation. EVT estimation requires a relatively large number of data points above a large loss threshold, further increasing the demand for empirical data points. The data scarcity may cause situations with the potential of having no losses in Event Type 4 (ET4) on a particular day while having multiple losses in ET7 on the same day. El-Gamal, Inanoglu, Stengel (2007) have suggested a solution to this "asynchronicity of loss arrival" problem by aggregating daily losses collected into weekly[15] losses (avoiding dependency problems for unmatched severity).

[Weekly aggregated loss data (LDCE 2008 will give approximately 350 weekly data items; LDCE 2004 gave 196 weekly observations). Hopefully, the larger number of observations in LDC32008 will improve fittings. Even for LDCE 2008, monthly data will probably provide too few observations for monthly estimation.]

Daily losses will be aggregated to weekly losses across loss event-types (ET) and not across business lines. Using business lines almost surely means combining data with *demonstrably different* risk profiles within the same loss distribution although the event-loss-types have very different characteristics and avoid the problem.

It may be interesting to retest rank-correlation estimates for annual losses in each unit of measure against rank-correlation estimates of observed weekly data. It may also be useful to implicitly test tail dependency measures by examining proportions of simulated losses in top 95% of each unit of measure occurring in simultaneous years and comparing the proportion in the original weekly data.

C. Choosing the Distributional Copula and Marginals

The copula chosen for the model determines the dependency properties of the joint distribution distinct from the choice of marginals. We will examine three distributional copula structures to maintain conservatism. Two of the chosen copulas allow consideration of positive tail dependence in addition to rank correlations. Positive tail dependence effectively results in reduced "diversification" since it implies large coincident losses occasionally occur in different units of measure. (Recall that negative tail dependency is infeasible for operational-risk loss distributions with one-sided support. Figures 3 and 4 have clearly shown that meta-t distributions with one-sided

[15] Giacometti, Rachev, Chernobai, and Bertocci (2008) reached a similar conclusion about the need to use weekly aggregation.

marginals do not exhibit the negative tail dependency responsible for diversification of large values.) Meta-t copulas do have multiple parameters to express correlation between different units of measure. Alternatively we will examine the (Archimedean) Gumbel copula which also exhibits positive tail dependence (see Figures 2 and 4 in Part II). Finally, we will contrast results with the Gaussian copula which exhibits no asymptotic tail dependency.

The choice of marginals to accompany the copula is vital for obtaining conservative results which do not inappropriately increase the measured impact of "diversification" and reduce required capital inappropriately. In a recent study, Giacometti et al (2008) used an alpha-stable distribution to estimate losses in three business lines for a single European bank and used t-copulas (including skewed-t) to represent dependence. Their results indicated an approximate 20-25% reduction in capital at the 99.9th percentile but strangely indicated superadditivity at some lower percentiles like the 98th. They wrote,

> "Our findings indicate a lower reduction in risk capital than those reported by Chapelle *et al* (2004) and Dalla Valle *et al* (2008). Chapelle *et al* (2004) showed that for a 99% VaR a bank can achieve reduction in capital estimates by 35% by using copulas, and Dalla Valle *et al* (2008) showed reduction in the range of 30% to 50%. One explanation is that we used a much more heavy-tailed loss distribution than in the other two studies. This means that, if a thin-tailed loss distribution is used to model loss data, copula-based aggregation would result in a much more significant reduction in total capital. In essence, this implies that if a bank has mistakenly chosen to use a thin-tailed loss distribution when the data is in fact heavy-tailed, the resulting capital would be understated. The converse is also true. The choice of the loss distribution thus becomes of central concern and must be estimated with a high degree of accuracy."

In this paper, we acknowledge the potential heavy-tailed nature of operational-risk losses by attempting to fit an EVT-mixture distribution to bank data. The loss-distributional marginals for various loss event types will be modeled as mixtures of lognormals with GPD tails. If the mixture fails to converge at a sufficiently low threshold to provide adequate data in the tail, a thin-tailed lognormal distribution will be assumed. The marginals will be fitted by MLE. Interestingly, even the application of MLE to estimate parameters of the marginal distributions requires conservative examination of an appropriate method to determine the threshold and to implement initial parameter-guesses for the underlying distributions. See section III.E for a discussion of methodological problems.

Since the mixture distribution (namely its tail) has the potential to provide an infinite mean model when considering excesses over a high threshold, the model provides the potential to create superadditive OpVaR for different units of measure as discussed in Part II. In such cases, clear outliers may be visible in the data. Potentially such outliers need to be cleared from the data used for MLE and modeled separately by scenario analysis.

In order to maintain a conservative approach, it is imperative to hypothesize and test distributional copulas which induce positive tail dependency suggesting the potential for large simultaneous losses in some units of measure. The analysis in Part II indicated that copulas with positive tail dependencies describe non-diversifiable losses in operational risk. Furthermore, the absence of negative tail dependence in one-sided operational-risk loss distributions may greatly reduce diversification among extreme operational-risk losses relative to losses observed in financial markets.

D. Choosing a Method to Fit a Distributional Copula

Three potential methods are currently used to fit a distributional copula to its marginals and determine the copula parameters.

1. **Method of maximum pseudolikelihood**

   The method of maximum pseudolikelihood is championed by Genest and Favre, 2007. It involves maximizing a rank-based, log-likelihood of the bivariate form by rescaled versions of their empirical distribution functions (dividing by n+1):

   $$\ell(\theta) = \sum_{i=1}^{n} \log\{c_\theta(\frac{R_i}{n+1}, \frac{S_i}{n+1})\}.$$

   Genest and Favre acknowledge that their exclusive advocacy of ranks for copula parameter estimation does not have complete consensus in the statistical community.

2. **Full MLE**

The copula and marginal parameters could be estimated simultaneously using full MLE. This is often extremely difficult and has been largely ignored. However, McNeil. Frey, Embrechts (2005) suggest

> "One could envisage using the two-stage method [IFM] to decide on the most appropriate copula family and then estimating all parameters (marginal and copula) in a final fully parametric round of estimation."

3. **IFM (Inference Functions for Marginals) Method**

The IFM is a two-step method first described in the textbook by Joe (1997).

Step 1: Fitting the Marginals under IFM

In the first step of IFM, parameters δ_i for each of the i marginal distributions may be estimated parametrically by MLE and a d-dimensional **pseudo-sample** of marginal probabilities can be built from the estimated marginals as

(4) $$\hat{U}_{t,1},...\hat{U}_{t,d} = (\hat{F}_1(X_{t,1};\delta_1),...\hat{F}_d(X_{t,d};\delta_d)).$$

Alternatively, the marginals may be estimated non-parametrically from the empirical distribution function (edf) although edfs "are known to be poor estimators of the underlying distribution in the tails"[16]. Furthermore, the pseudo-sample marginals generated by edf will have identical marginals in each column (but with different ordering), requiring not only independence but identical marginals across all loss types. Consequently, the non-parametric edf approach is not recommended for fitting marginals for operational-risk data with potentially heavy tails and different distributions in each unit of measure. As a third alternative, one could try to fit the tail parametrically[17] and the body non-parametrically.

In section III.E we present details for implementing the first IFM step via MLE using a lognormal-GPD mixture distribution. Note that dramatically different results can occur depending upon how initial guesses are implemented as inputs to MLE and how the threshold level is empirically determined when the data is heavy-tailed.

Step 2: Fitting the Copula Parameters under IFM using MLE

In the second step of IFM, copula parameters θ are estimated by MLE conditional upon the first-step estimators of the marginals-distribution parameters δ_i. Note that for meta-copulas, each marginal may have a different set of parameters from the others. The loglikelihood function is written in terms of the copula density c_θ with respect to the copula parameter vector θ

(5) $$\max_{\{\theta\}} \ln L(\theta;\hat{U}_{1,t},...\hat{U}_{d,t};\delta_1,...\delta_d) = \sum_{t=1}^{n} \ln c_\theta(\hat{\mathbf{U}}_t;\boldsymbol{\delta}).$$

Hence in order to fit copula parameters under Step 2 of IFM, the implementer must determine the copula's density function which is not always a trivial task since implicit copulas have no closed-form solutions and explicit copulas may be very tedious.

Obviously, the statistical quality of the copula parameter(s) fitted in step 2 of IFM is strongly impacted by the accuracy of the estimators for the marginal distributions in Step 1. Nevertheless, IFM is strongly preferred over full MLE by most practitioners because it is easier to implement. Xu (1996) and Joe (1997) both studied the relative efficiency of the IFM estimators versus the full MLE and concluded that the IFM method is highly efficient with mean-squared error ratio to the full MLE close to 1.[18]

See Section III.F for details of implementing the second IFM step, including determining the copula density functions needed for likelihood estimation of the copula parameters.

---

[16] See McNeil, Frey, Embrechts (2005) text, p.233.

[17] For example, the tail could be fitted parametrically as a GPD. There are numerous methods other than MLE to estimate the shape and scale parameters of the GPD. For example, Juarez and Schucany developed the minimum density power divergence estimator (MDPDE) while Dupuis developed the optimally biased robust estimator (OBRE) for parameters in a GPD tail.

[18] See Bee (2005). For Joe, see section 10.1.2.

This project will simulate a large set of joint losses from three different distributional copulas (Gaussian, t, and Gumbel) using a hybrid of the two-step IFM method. In the second step, copula parameters for a standard copula will be determined by non-parametric estimation using the empirical distribution function. Important statistics like VaR and the diversification (capital reduction) ratio will be determined from the empirical distribution of the simulated loss set. See Section III.G for this hybrid IFM approach useful when simulating data from meta-copulas.

Details follow in the next three sections (E,F,G). Note that in all cases, the underlying operational-risk loss marginals have one-sided support ($X \in (0,\infty)$) which precludes the major diversification benefits of negative tail dependency available from two-sided support available in the capital markets.

E. Fitting the Marginals via MLE: Step 1 of IFM

The first Inference Functions for Marginals (IFM) step estimates the marginal loss distributions for each unit of measure (here assumed to be event-types ET) in an individual bank. We initially show diversification results for a single large U.S. bank whose operational-risk losses were provide in the LDCE2004 dataset. [We plan to extend results by separately fitting and comparing several banks with sufficient operational-risk loss data provided in the LDCE 2008 dataset.] In Part III.C we indicated that assumed marginals would be either lognormal or a mixture of lognormal with a GPD tail. If data-fitting fails to support a lognormal-GPD mixture for an ET, then the marginal for that unit of measure will be fitted to a straight "medium-tailed" lognormal distribution. We will fit the marginals using maximum likelihood by simultaneously estimating parameters for both the body and the tail.

The appropriate mixture distribution (and corresponding density) can be derived from Extreme Value Theory (EVT). Given the Generalize Pareto Distribution, the basic formula to **relate** the **tail probabilities** (the "excess" conditional distribution function) to the **full loss distribution** is given by the **conditional excess distribution $F_u(x)$** in the following formula where *X represents the actual value observed* (an *exceedance* over the threshold) while x represents the "excess" value over the threshold (x = X-u). Hence in the following notation, *X* is an *exceedance* and *x* is an *excess*. The following equation describes the *distribution of an excess* value x where the "excess" measures the difference between the actual "exceedance" observation X and the threshold u.

$$(6) \quad F_u(x) \equiv F_u(X-u \mid X>u) = \frac{F(u+x)-F(u)}{1-F(u)} = \frac{F(X)-F(u)}{1-F(u)} = \frac{\textit{area between } X \textit{ and threshold } u}{\textit{total area above threshold } u}$$

Recognizing that F(X) is a lognormal distribution and changing notation so that the threshold becomes $T_\lambda = u$ where λ is the percentile associated with the level T = u and identifying that $F_u(x) = F_u(X-T_\lambda) = F_{GPD}(X-T_\lambda;\xi,\beta)$, we find that the aggregate loss distribution function may be rewritten as a *mixture* of a lognormal distribution $F_{LN}(X;\mu,\sigma)$ *if* $X \le T_\lambda$ below a threshold and a GPD excess distribution function above the threshold where ξ is the shape parameter and β is the scale parameter of the GPD:

$$(7) \quad F(X;\mu,\sigma,\xi,\beta,T_\lambda) = \left\{ \begin{array}{ll} F_{LN}(X;\mu,\sigma) & \textit{if } X \le T_\lambda \\ (1-F_{LN}(T_\lambda))*F_{GPD}(X-T_\lambda;\xi,\beta)+F_{LN}(T_\lambda;\mu,\sigma) & \textit{if } X > T_\lambda \end{array} \right\}$$

Differentiating the distribution function gives the mixture distribution's density (likelihood function) which may be converted to logarithms and used to simultaneously estimate the parameters for both the lognormal and the GPD tail via MLE:

$$(8) \quad L(X;\mu,\sigma,\xi,\beta,T_\lambda) = \begin{array}{ll} f_{LN}(X;\mu,\sigma) & \textit{if } X \le T_\lambda \\ (1-F_{LN}(T_\lambda;\mu,\sigma))*f_{GPD}(X-T_\lambda;\xi,\beta) & \textit{if } X > T_\lambda \end{array}$$

To maximize the log of the likelihood function in equation (8), we fix a set of threshold percentiles $\lambda = \{\lambda_0, \lambda_0+\varepsilon, \lambda_0+2\varepsilon, ..., \lambda_0+n\varepsilon\}$ where ε is a very small number like 0.005. For each percentile in the set, we evaluate the loglikelihood with respect to the other four parameters $\{\mu,\sigma,\xi,\beta\}$ at the implied

threshold T_λ. From the resulting collection of loglikelihood values at each threshold, we select the value which minimizes the negative loglikelihood function to define the optimal threshold.

To promote conservatism, it is important to examine different modeling approaches and even different software implementations[19] for MLE which may result in significant differences in parameter estimates for the marginal distributions. We will show that at least three factors can significantly impact estimations for "heavy-tailed" data when fitting a mixture or GPD distribution to historical data.

i. Judicious removal of outliers from loss event-type data when initial analysis indicates "extremely-heavy" tails (infinite-mean distributions).
ii. The method chosen to **initialize** parameter guesses when applying MLE to heavy-tailed distributions (using either PWM or MOM[20] statistics as starting values).
iii. Choice of potential threshold percentile ranges within the MLE procedure and differences in the empirical quantile function used to obtain the actual threshold from a specified threshold percentile.

1. Outliers Factor

Several analysts have suggested that outlier removal may be justified when the MLE estimators for a loss event-type imply an infinite-mean distribution[21]. For example Neslehova et al (2006) state

> "VAR may lead to ridiculously high capital charges in the infinite-mean case…. One could even argue that infinite-mean models ought to be banned from the operational risk modelers toolkit!"

The removal of outliers from an event-type will clearly diminish parameter estimates for the shape (and possibly scale) parameters, resulting in a reduction in OpVaR and consequently required capital.

El-Gamal, Inanoglu, Stengel (2007) suggest a strategy of judiciously re-estimating the data series after "separating clear outliers that need to be modeled via probabilistic scenario analysis". When removing outliers and reducing OpVaR, there should be an offsetting procedure to adjust the OpVaR upward via scenario analysis to account for the removed outliers. The authors state that Bayesian analysis should be applied to the outliers and outlined an ad hoc procedure for adjusting the OpVar upward to account for omitted outliers. However, they reiterate that the "full treatment of this procedure is subject to future research", echoing the necessary conservatism message of the NPR requiring future improvements to our understanding of the issue.

Table 1. Impact of Excluding Outliers

Loss Type	Outliers Excluded	Negloglik	Threshold Value	Initializer Type	xi (GPD shape)	beta (GPD scale)	logmu	logsd
ET4	0	3019.483	1,606,367 (68.5%)	PWM7	1.1885 (0.3079)	6.093835e+06 (2.5925e+06)	13.772 (8.316e-02)	1.0832 (7.1109e-02)
ET4	4	3005.077	1,609,515 (70%)	PMW7	0.7788 (0.27179)	4.278366e+06 (1.3145e+06)	13.741 (8.2328e-02)	1.0683 (6.993e-02)
ET2	0	3095.093	3,737,860 (63.5%)	PWM7	0.48568 (0.14622)	1.976857e+06 (3.211684e+05)	15.0172 (0.02547)	0.32377 (0.0222)
ET2	4	3030.369	NA	PWM7	NA	NA	15.1147 (0.034074)	0.47215 (0.02409)

Tables 1-3 report the impact of outliers on parameter estimations for the marginal distributions of data from a single large U.S. bank as reported in LDCE 2004 data. Table 1 shows the impact on shape and scale parameters for

[19] By different implementations, we mean using Matlab, R, S-Plus, or other packaged functions.

[20] Probability Weighted Moments are quantiles $E[X^p F(X)^r (1-F(X))^s] = M_{p,r.s}$. For the GPD, we usually work with $M_{1,0,s}$, i.e. setting p=1, r=0, s=s. MOM indicates method of moments. See Hosking and Wallace (1987), section 3.3 especially equation (5).

[21] An extreme value distribution has an infinite variance if its shape parameter chi is greater or equal to 0.5 ($\xi >= 0.5$) and an infinite mean if its shape parameter is greater or equal to 1.0 ($\xi >= 1.0$).

two loss event-types when the four largest outcomes (outliers) are removed from ET4[22]. The first row of Table 1 shows the MLE estimators for ET4's parameters $\{\xi, \beta, \mu, \sigma\}$ using the full data set with initial guesses calculated via probability-weighted-moments (PWM) and using the mode-type empirical quantile function (Type 7) for the entire data set of 196 aggregated weekly losses[23]. The outliers were removed because the ξ-estimator of 1.1885 for ET4's shape parameter with the full dataset indicates an infinite-mean model ("extremely-heavy-tailed" distribution) which could result in superadditivity. Hence the outlier-reduced data set has 192 weekly observations. Obviously, the corresponding weekly observations must also be removed from all other loss types.

The second row of Table 1 shows revised parameter estimates for ET4 after removing the four outliers. Note the ξ-estimator of 0.7788 now indicates an acceptable infinite-variance model ("heavy-tailed" distribution) no longer susceptible to superadditivity.

It is obvious that the removal of four items from each dataset may at least marginally impact resulting parameter estimation for all other loss event types as well.

In fact, examination of rows 3 and 4 of Table 1 indicates a decline in the shape parameter of ET2 which is so steep that ET2 can now be estimated as a straight lognormal distribution with a narrower tail than a mixture. How do we explain this? Ex-post scrutiny of the outliers shows that removal of the four largest values from ET4 also resulted in removal of the largest, 21st largest, 23rd largest, and 128th largest items from ET2, basically destroying its tail. (Additionally, the 2nd largest loss was removed from ET1.) Hence ET2 and ET4 clearly displayed strong positive tail dependence whose effect was diluted by removing the outliers. This undesirable consequence must also somehow be modeled by scenario analysis and is an example of the need for future improvements in our modeling abilities.

Table 2 shows the estimators for modeled event types ET1, ET2, ET3, ET4, ET7 for the full dataset of 196 observations. (The bank did not have sufficient data to properly aggregate ET5 and ET6 on a weekly basis.) Table 3 shows the corresponding adjusted parameter estimators after removal of the outliers. The results in Tables 2 and 3 will serve as input parameters for the quantile functions needed to invert probabilities simulated from copulas in Part III and obtain simulated losses for calculating the diversification ratios for various copulas.

Table 2: Parameter Estimates for All Loss Types With Full Data (PW7)

Loss Type	Negloglik	Threshold Value	xi (GPD shape)	beta (GPD scale)	logmu	logsd
ET1	2719.886	397,560 (76.5%)	0.8822 (0.2517)	4.694734e+05 (1.22301e+05)	12.347 (6.168e-02)	0.8283 (5.039e-02)
ET2	3095.093	3,737,860 (63.5%)	0.48568 (0.14622)	1.976857e+06 (3.211684e+05)	15.0172 (0.02547)	0.32377 (0.0222)
ET3 (fit as LN)	2834.664	NA	NA	NA	13.1458 (0.06448)	0.90283 (0.04560)
ET4	3110.117	1,606,367 (68.5%)	1.18852 (0.3079)	6.093835e+06 (2.592499e+06)	13.7721 (0.008316)	1.08321 (0.007111)
ET7	2942.212	1,301,337 (28%)	0.40001 (0.10322)	5.65203e+05 (7.289927e+04)	14.5031 (0.00932)	0.71070 (0.00786)

Further comparison of Tables 2 and 3 shows that the removal of outliers induced only minor differences for ET1, ET3, and ET7 parameters. ET1 remained a "heavy-tailed" (infinite-variance) distribution (although the removal of the 2nd largest observation as an outlier undoubtedly explains the reduction in the shape parameter). ET3 remained open to fitting a straight lognormal rather than a mixture, and ET7 remained a "medium-tailed" distribution with a

[22] Each removed outcome exceeded seven standard deviations from the mean. The seven event loss types defined by Basel II include ET1 (Internal Fraud), ET2 (External Fraud), ET3 (Employment Practices and Workplace Safety), ET4 (Clients, Products, and Business Practices), ET5 (Damage to Physical Assets), ET6 (Business Disruption and System Failures), ET7 (Execution, Delivery, and Process Management).

[23] Review section III.B for an explanation of weekly aggregation.

fairly high (but finite) variance. We have already discussed the **major** change for ET2 induced by the outlier removal.

Table 3: Parameter Estimates for All Loss Types Outliers Removed (PW7)

Loss Type	Negloglik	Threshold Value	xi (GPD shape)	beta (GPD scale)	logmu	logsd
ET1	2655.966	396,028 (76.5%)	0.75957 (0.23883)	4.193910e+05 (1.020006e+05)	12.34447 (0.06172)	0.82052 (0.00504)
ET2 fit as LN	3030.369	NA	NA	NA	15.1147 (0.034074)	0.47215 (0.02409)
ET3 fit as LN	2763.51	NA	NA	NA	13.1162 (0.06262)	0.8677 (0.04428)
ET4	3005.077	1,609,515 (70%)	0.78738 (0.27179)	4.278366e+06 (1.3145e+06)	13.741 (8.2328e-02)	1.0683 (6.993e-02)
ET7	2873.552	1,375,373 (33.5%)	0.45726 (0.11516)	4.901010e+05 (6.763896e+04)	14.307 (0.007535)	0.65199 (0.006566)

Our analysis has shown that to maintain conservatism, it is important to clearly note the impact of outlier removal on the other datasets as well. Besides impacting ET4's marginal distribution, the removal of four outliers from event type ET4 also may have significantly reduced positive **tail dependence** among event types ET4 and ET2 since the removed weeks had several simultaneous extremely large losses. Hence the outlier removal affects not only the marginal distributions for ET2 and ET4 in this section but also tail dependence reflected in ensuing copula estimation in part III. There are no existing models to handle this change explicitly. Possibly some increase in capital charge could be implemented through scenario-analysis. The example accentuates the need for conservatism.

2. Initial-Guess Factor for the MLE Algorithm

Our study indicates a second somewhat unexpected factor also appears to significantly influence the size of the GPD shape and scale parameter estimates. Existing software uses several different methods to input initial guesses into the loglikelihood function. In an early study fitting the generalized Pareto distribution (GPD) via MLE, Hosking and Wallace[24] (1987) described alternative methods to MLE including PWM (probability-weighted-moment) and MOM (method-of-moment) estimators. They also described a standard procedure of initializing guesses[25] for the MLE estimation by using either PMW- or MOM-initializers. Their study analyzed only right-truncated, “thin-tailed”, or “medium-tailed” distributions with shape parameter less than one-half, implying that mean and variance are always finite. They reported that MLE frequently failed to converge under all initializers when analyzing hydrological data unless the sample sizes were very large. Today, other available computer code also initializes MLE parameters using either the Pickands estimators or the Coles estimators.[26] However, much current MLE code seems to be gravitating toward MOM- and PWM-initializers. initialized via MOM. Nevertheless,

[24] The Hosking-Wallace study largely dismissed cases which are now known as "heavy-tailed" by claiming that distributions with infinite variance are "unusual in statistical applications". Since 1987, financial literature has become entirely accustomed to return data with havy tails implying infinite variance. Unfortunately, the operational-risk literature today is also filled with examples of data fitting results (estimators) implying "extremely heavy-tailed" distributions characterized by infinite means. Many statisticians believe that such estimators are generally the result of data contamination, i.e. inappropriately mixing data from distributions with different risk characteristics.

[25] Hosking and Wallace noted that their tests initialized the Newton-Raphson method “from a variety of starting values of a and k [scale and shape parameters], some based on the sample moment’s or PWMs, some chosen at random”. Following Hosking and Wallace, we examine only the MOM and PWM approaches to initialization. Less known (more esoteric) approaches to initialize guesses (like the Pickand's function and the Cole function) are also available in add-ons to the S-Plus software. See the evanesce library for S-Plus.

[26] For example, Juarez and Schucany have published S-Plus software which use Pickands or Cole estimators as initializers in the gpd.mle() method

Table 4 shows contrasting R-language computer code for generating the initializers under both MOM and PWM methods. The PWM code shown is consistent with Alexander McNeil's newest implementation of QRMlib software (both S-Plus and R editions). The new fExtremes R-language package from RMetrics also initializes the MLE estimator using a PWM. McNeil currently recommends using the PWM method in all cases where the fitted underlying distribution being might be either "heavy-tailed" (having an infinite variance, i.e. a shape parameter exceeding 0.5) or "extremely-heavy-tailed" (having an infinite mean, i.e. a shape parameter exceeding 1.0). He asserts that accurate MOM estimators for starting guesses *cannot* be obtained whenever the underlying distribution has any infinite moments since all MOM parameters are finite by assumption. This contradiction clearly could

Table 4. R-language Computer Code for PWM and MOM Initializers

MOM Initializers Code

```
initialParameterGuessesMOM <- function(lData, threshPercentile, qType)
{
    #R-default is type=7.  Matlab default is type=5 which always results in a threshold HIGHER than the
    #default R case with type=7.
    thresh <- quantile(lData, threshPercentile,type=qType); #(default is type=7; Matlab uses type=5);
    #Divide the data set into two parts, body and tail and set the body as the data less or equal to the threshold
    bodyData <- lData[lData <= thresh];
    #set the tail as the data greater than the threshold
    tailData <- lData[lData > thresh];
    #Set the initial parameters using MOM methods:
    #first two for LN:
    d1 <- mean(log(bodyData));
    d2 <- abs(sd(log(bodyData)));
    #Second two for GPD:
    m3 <- mean(tailData-thresh);
    s3 <- abs(sd(tailData-thresh));
    d3 <- -((m3^2)/(s3^2)-1.0)/2.0;
    d4 <- m3*((m3^2)/(s3^2)+1.0)/2.0;
    out <- c(d3,d4,d1,d2);
    return(out);
 }
```

PWM Initializers Code

```
initialParameterGuessesMOM <- function(lData, threshPercentile,qType)
{
    thresh <- quantile(lData, threshPercentile,type=qType);
    #Divide the data set into two parts, body and tail and set the body as the data less or equal to the threshold
    bodyData <- lData[lData <= thresh];
    #set the tail as the data greater than the threshold
    tailData <- lData[lData > thresh];
    #Set the initial parameters using MOM methods for LN:
    #first two for LN:
    d1 <- mean(log(bodyData));
    d2 <- abs(sd(log(bodyData)));
    #Set second two for GPD according to PWM (probability-weighted moments) method described in 1987 paper by
    #Hosking and Wallace:
    #"Parameter and Quantile Estimation for the Generalized Pareto Distribution"
    excesses <- tailData - thresh;
    Nu <- length(excesses);
    a0 <- mean(excesses);  #this is alphasub0 in Hosking and Wallace
    gamma <- -0.35; #these constants (gamma, delta) were recommended by Landwehr according to Hosking and Wallace
    delta <- 0;
    pvec <- ((1:Nu) + delta)/(Nu + delta);
    a1 <- mean(sort(excesses) * (1 - pvec));
    xi <- 2 - a0/(a0 - 2 * a1);   #note this is k in equation (5) from Hosking and Wallace
    beta <- (2 * a0 * a1)/(a0 - 2 * a1);  #Note this is alpha in equation (5) from Hosking and Wallace
    out <- c(xi,beta,d1,d2);
    return(out);
 }
```

have a significant impact on any resulting MLE estimators initialized via MOM. Nevertheless, other software versions still initialize MLE estimators using MOM. For example, El Gamal, Inanoglu, Stengel (2007) used

software with MOM-initializers for the MLE process in their paper demonstrating the need for judicious use of outliers.

Indeed, Table 5 shows large discrepancies between MLE estimators using PWM- and MOM-initializers. The table also shows discrepancies induced by the use of alternative empirical quantile functions (type= 5 or type=7) to determine the threshold where the GPD tail starts.

Table 5.
MLE: Can PWM Initializers Properly Identify "Thin-Tailed" Distributions? Can MOM Initializers Properly Identify "Heavy-Tailed" Distributions?

Using Full Data Set (No Outliers Excluded)

Loss Type	Negloglik	Threshold Value	Initializer Type	xi (GPD shape)	xi initial guess	beta (GPD scale)	logmu	logsd
ET3	2822.579	1,806,880 (96%)	PWM7	0	-0.2213	5.842719e+06	13.112	0.8188
ET3	2823.526	1,806,880 (96%)	MOM7	0	-0.7746	8.489172e+06	13.112	0.8188
ET3	2824.033	1,734,360 (95.5%)	MOM5	0	-0.3898	5.999949e+06	13.112	0.8196
ET1	2719.886	397,560 (76.5%)	PWM7	0.8822 (0.2517)	0.8158	4.694734e+05 (1.22301e+05)	12.347 (6.168e-02)	0.8283 (5.039e-02
ET1	2723.295	2,057,329 (97.5%)	MOM7	0	0.02957	1.699802e+07	12.434	0.9698
ET1	2725.168	1,893,601 (97%)	MOM5	0.2485	0.1526	1.249005e+07	12.437	0.9673

Specifically, Table 5 shows MLE estimators for all four parameters $\{\xi, \beta, \mu, \sigma\}$ of the lognormal-GPD mixture distribution in (7) for two event loss types (ET1 and ET3) from a single bank as reported in LDCE 2004 data. Parameters are derived by using the entire data set of 196 aggregated weekly losses[27] . The optimal threshold value T_λ is obtained from one of two different empirical quantile functions {Threshold Type5, Threshold Type7}. Initial parameter guesses are calculated from one of two methods {PWM, MOM}. Hence the "Initializer-Type" column reflects a combination of parameter initializer method with empirical quantile method. For example, PWM7 indicates a choice of PWM initializers combined with a Type 7 (R-language default) selection for the empirical quantile function whereas MOM5 indicates method-of-moments initializers and a Type 5 (Matlab default) empirical quantile function.

Note that for ET3 losses, both the PWM- and MOM-initializers imply a similar "medium-tailed" distribution since the shape parameter is approximately zero regardless of the quantile method or initializer-method selected.

However, the PWM and MOM methods provide drastically different results for losses in ET1. The PWM7 method for ET1 implies a "heavy-tailed" distribution with *infinite variance* (since the estimator $\xi = 0.8822 > 0.5$). However, the MOM7 and MOM5-initializer methods both imply a finite variance model with the MOM7 method suggesting a shape parameter approximately zero $\xi \cong 0)$ while the MOM5 method suggests a shape parameter of $\xi \cong 0.2485$. Hence MOM again induces a discrepancy consistent with requiring finite variances even in cases where the underlying distribution may actually have infinite variance as suggested by the PWM7 estimator.

Table 6 shows how the initial guesses for the shape (ξ) parameter vary as threshold-percentiles increase for both event types ET1 and ET3. The second column shows that PW7 initial guesses for ET1's shape parameter never approach zero at any percentile, indicating at least a "heavy-tailed" distribution. In contrast, the fourth column shows that the PWM7-initial guesses (calculated by the formulas shown in Table 4) will finally approach zero as the specified threshold percentile approaches 97.5%. Hence the PWM7 method is able to handle all types of distributions: those whose initial-guesses never approach zero (implying "heavy-tailed" or "extremely-heavy-tailed"

[27] Review section III.B for an explanation of weekly aggregation.

distributions) and those whose initial-guesses approach zero (implying "medium-tailed" or "thin-tailed" distributions).

Note that the MOM initializers are dramatically smaller than the PWM initializers for the ET1 distribution. Since the MOM initializers by definition have finite moments, it is difficult for MOM to provide large enough initial guesses to describe either a "heavy-tailed" (infinite-variance) or "extremely-heavy-tailed"(infinite-mean) underlying distributions like ET1.

Table 6. MLE: PWM vs MOM Initial-ξ Guesses at Various Percentiles

Using Full Data Set (No Outliers Excluded)

Percentile	ET1 Initial ξ-Guess PWM7 Implies "heavy tailed" Optimal ξ =0.8822	ET1 Initial ξ-Guess MOM7 Implies "medium/thin tailed" Optimal ξ =0	ET3 Initial ξ-Guess PWM7 Implies "medium/thin tailed" Optimal ξ =0	ET3 Initial ξ-Guess MOM7 implies "medium/thin tailed" Optimal ξ =0
0.2	0.8157	0.4789	0.4311	0.3736
0.205	0.8183	0.4790	0.4311	0.3736
0.21	0.8186	0.4789	0.4338	0.3740
…				
0.445	0.8438	0.4742	0.5395	0.3831
0.45	0.8436	0.4740	0.5340	0.3819
0.455	0.8432	0.4738	0.5345	0.3816
…				
0.695	0.8370	0.4594	0.6539	0.3743
0.7	0.8391	0.4590	0.6521	0.3743
…				
0.76	0.8176	0.4498	0.6295	0.3521
0.765	0.8158**	0.4488	0.6262	0.3498
…				
0.9	0.8519	0.4098	0.6069	0.2478
0.905	0.8509	0.4057	0.5952	0.2347
…				
0.955	0.7741	0.2963	0.0253	-0.3917
0.96	0.7409	0.2617	-0.2213**	-0.7746**
0.965	0.7035	0.2162	-0.4828	-1.3822
0.97	0.6563	0.1502	-0.6812	-2.2716
0.975	0.5732	0.0295**	-0.3139	-1.3569
0.98	0.4915	-0.1642	-0.9761	-14.0563
0.985	0.3796	-0.6751	-0.0377	-2.7572
0.99	0.3309	-5.8606	0.9139	0.1469
** Indicates Initial Guess at Optimal Percentile				

Table 7 shows that a MOM-initializer actually can imply an infinite-variance distribution. Row 2 shows results

Table 7. Implied Infinite-Mean Model Fit Via PWM/ MOM Initializers

Using Full Data Set from Existing Bank (No Outliers Excluded)

Loss Type	Negloglik	Empirical Threshold Type	Threshold Value	Initializer Type	xi (GPD shape)	xi initial guess	beta (GPD scale)	logmu	logsd
ET4	3110.117	Mode (type7)	1,606,367 (68.5%)	PWM	1.1885 (0.3079)	0.8545	6.093835e+06 (2.5925e+06)	13.772 (8.316e-02)	1.0832 (7.1109e-02)
ET4	3120.011	Mode (type7)	4,222,013 (83.5)	MOM	0.72720	0.39249	4.562920e+07	13.892	1.248

from MOM7 initial guesses which imply a shape parameter of ξ = 0.7272 for ET4. However, row 1 in Table 7 show that the PWM7-initializer indicates an infinite-mean ("extremely heavy-tailed") distribution with a shape parameter estimator of ξ = 1.1885. Hence the MOM-initialized MLE estimator for the shape parameter value is

substantially below the value suggested from PWM-initializers while the opposite holds for the scale parameter where the MOM initializer is substantially higher than that forecast by PWM.

The results in Tables 5-7 fitting unknown distributions cast doubt on the MOM-initializer method, implying that it may significantly underestimate the shape parameter (and over-estimate the scale parameter) in all “heavy-tailed” distributions, although it performs satisfactorily for underlying “medium-tailed” or “thin-tailed” distributions which have finite variances.

Further evidence that MOM-initializers routinely cause a significant underestimation of the shape parameter can be found by applying MLE to simulated data with known parameter values. Results differ slightly with the number of simulated values. In one case, suppose we simulate 1000 values from a lognormal-GPD mixture using equation (7) with an infinite-mean ("extremely heavy") tail ($\xi > 1$); in a second case, we simulate only 196 points. Specifically we utilize the parameters shown in the following table (consistent with the first row of Table 7):

paramMatrix for simET4

threshold	xi	beta	meanlog	sdlog
1.606367e+06	1.188518	6.093835e+06	13.7721	1.083211

We then fit the simulated data via MLE. Table 8 contrasts the resulting fits obtained via the PWM7 and MOM7 methods for choosing initializers and empirical quantile functions with both simulations.

Table 8. Simulated Infinite-Mean Model

Simulated Distribution with $\xi = 1.1885$

Loss Type	#Simulated Data Points	Negloglik	Threshold Value	Initializer Type	xi (GPD shape)	beta (GPD scale)	logmu	logsd
SimET4	196	3130.477	1,603,709 (68.5%)	PMW7	0.88075 (0.25879)	9.308351e+06 (2.77808e+06)	13.744 (0.0889)	1.1597 (0.07608)
SimET4	196	3136.409	1,603,709 (68.5%)	MOM7	0.580296	1.771175e+07	13.751	1.1592
SimET4	1000	15866.63	1,543,857 (69.5%)	PWM7	1.015509 (0.1183)	8.473832e+06 (1.16553e+06)	13.6844 (0.0361)	1.0678 (0.03074)
SimET4	1000	15939.87	1,594,633 (70%)	MOM7	0.651099	2.507295e+07	13.6907	1.0756

Note that when fitting data from the large-sample simulation (1000 points), the PWM initializer attains a shape estimator of 1.0155 implying an infinite mean distribution whereas the MOM initializer implies only an infinite variance model via its shape-estimator of 0.65. When fitting the small simulation sample (196 points), neither initializer achieves infinite-mean estimator but the PWM estimator is much larger (closer) than the MOM estimator. This suggests that in small samples, even MLE with PWM7-initializers may slightly underestimate the shape parameter. The results in Table 8 are parallel to those in Table 7 where the MOM estimator significantly underestimated the shape parameter and produced an estimator consistent with an infinite-variance model rather than the appropriate infinite-mean model.

Alternatively, suppose we simulate values from a lognormal-mixture using equation (7) with an infinite-variance ("heavy") tail; we utilize the parameters shown in the following table (consistent with the first row of Table 2):

paramMatrix for SimET1

threshold	xi	beta	meanlog	sdlog
3.975600e+05	0.8822184	4.694734e+05	12.3472	0.82835

Once again we fit the simulated data via MLE. Table 9 contrasts the resulting fits obtained via the PWM7 and MOM7 methods for choosing initializers and empirical quantile functions in the infinite variance model. Note the results in Table 9 are parallel to those in Table 8. Here the MOM estimator significantly underestimates the shape parameter and falls significantly below the PWM estimator but does produce a value consistent with an infinite-variance model with $\xi > 0.5$ for the large simulation sample. However, for the small simulation sample, the MOM estimator is not even consistent with a "heavy" tail since the estimator is 0.268.

Table 9. Simulated Infinite-Variance Model Fit Via PWM/ MOM Initializers

Loss Type	#Simulated Data Points	Negloglik	Threshold Value	Initializer Type	xi (GPD shape)	beta (GPD scale)	logmu	logsd
SimET1	196	2727.098	379,271 (73%)	PWM7	0.64253 (0.25681)	5.668518e+05 (1.6337e+05)	12.3265 (0.06625)	0.87918 (0.0552)
SimET1	196	2727.729	599,347 (83.5%)	MOM7	0.26878 (0.2611)	1.292559e+06 (4.30335e+05)	12.4641 (0.06809)	0.93252 (0.0539)
SimET1	1000	13832.74	430,384 (78%)	PWM7	0.71041 (0.1158)	6.068967e+05 (7.8709e+04)	12.3000 (0.02763)	0.8436 (0.0225)
SimET1	1000	13842.36	430,384 (78%)	MOM7	0.51127 (0.10891)	9.592441e+05 (1.60233e+05)	12.3005 (0.02763)	0.8436 (0.0225)

Due to the consistent pattern of MOM estimators just described, we recommend for conservatism the use of PWM initializers for the MLE procedure in all cases where heavy-tailed distributions are suspected.

Finally, comparing Table 10A to 10C (or comparing Table 10B to 10D) shows that in all cases the MLE estimators of the GPD tail shape and scale parameters are substantially higher whenever the PWM method is used to initialize parameter guesses rather than when the MOM method is used.

3. Threshold Factor: Identifying Threshold Via Empirical Quantile Function

Numerous papers have specified that identifying the threshold (where the tail function begins) is one of the most important and yet frequently unresolved issues in applying extreme value theory to operational risk loses . In estimating a mixture distribution, we have chosen to specify a set of threshold percentiles $\{\lambda\}$, evaluate the negative loglikelihood function over the parameter set $\{\xi, \beta, \mu, \sigma\}$ at each potential threshold T_λ in the set, and then select the threshold corresponding to the parameter set which minimizes the resulting set of negative loglikelihood functions.

In our methodology, the selected percentile λ must be converted to a threshold value T_λ via an empirical quantile function. Different software packages like R, S-Plus, Matlab offer numerous versions of the empirical quantile function. For example, R and S-Plus have nine different choices set by one of the function's input parameters. The choice of the "type" parameter for the standard empirical quantile function[28] influences the threshold value and hence the resulting MLE estimators. Not all languages use the same default mode for the empirical quantile function so it is important to determine which method your software is using. For example, the default mode in the R-language basically represents the mode of the distribution function. (It is specified as Type=7 in R-language code). A popular alternative among hydrologists is a **piecewise linear function** where the knots are the values midway through the steps of the empirical cdf; this is the default in Matlab and may be used by specifying Type=5 in R-language code. The alternative hydrologic empirical quantile function always gives a higher threshold than does the default "mode" function, potentially resulting in too few observations above the threshold for convergence.

A comparison of combinations of PWM initializers with (Type5, Type7) is shown in Tables 10C and 10D for three different event loss types (ET2, ET4, ET7) on the data set with 192 weekly observations created by excluding four outlier weeks from all data sets. A similar comparison of MOM initializers is shown in Tables 10A and 10B. Note once again that regardless of the threshold method, the shape parameter determined from PWM initializers exceeds the corresponding shape parameter from MOM initializers while the scale parameter follows the opposite pattern. However, the difference in scale parameters appears to be relatively small for medium-tailed datasets (with finite variances, $\xi < 0.5$) but appears to be substantial for heavy-tailed datasets (with infinite variances $\xi > 0.5$). This is consistent with the notion that MOM estimators are inapplicable when variances are infinite.

By comparing tables 10C and 10D, it appears that the threshold is always higher for PWM initial guesses when using the Type=5 empirical quantile function rather than the Type = 7 function. However, the same conclusion

[28] The S-Plus or R function is quantile(x, probs, type,…).

Table 10.
Differences in MLE Induced by Alternative Initialization Methods or Empirical Quantile Functions
Data with Outliers Removed

Table 10A. Using MOM initial guesses in MLE with Type 7 empirical quantile functions to set threshold.
In this case *p(k) = (k - 1) / (n - 1) = mode[F(x[k])]*.

Asset	Negloglik	Threshold (mode Type=7)	xi Estimator	xi initial guess (MOM7)	beta Estimator	logmu Estimator	logsd Estimator
ET2	3020.103	2,559,642 (20%)	0.1572393 (1.151216e-01)	0.122677	1.847599e+06 (2.688922e+05)	1.520446e+01 (1.657602e-01)	5.000000e-01 (1.51185e-01)
ET4	3006.611	2,438,308 (77.5%)	0.4135227 (2.530596e-01)	0.28837	9.821287e+06 (3.448313e+06)	1.379072e+01 (8.466759e-02)	1.131640 (6.933977e-02)
ET7	2874.064	1,375,373 (33.5%)	0.3928220 (1.088510e-01)	0.380642	5.766151e+05 (8.437442e+04)	1.443065e+01 (7.534158e-02)	6.520020e-01 (6.56618e-02)

Table 10B. Using MOM initial guesses in MLE with Type 5 empirical quantile functions to set threshold.
Here *p(k) = (k - 0.5) / n* is a **piecewise linear function** popular among
hydrologists where the knots are the values midway through the steps of the empirical cdf.

Asset	Negloglik	Threshold (Piecewise Linear) Type=5	xi Estimator	xi initial guess (MOM5)	beta Estimator	logmu Estimator	logsd Estimator
ET2	3019.483	2,555,575 (20%)	0.1610495 (1.153478e-01)	0.12550	1.833252e+06 (2.674535e+05)	1.521024e+01 (1.622711e-01)	5.000000e-01 (1.452744e-01)
ET4	3006.478	2,704,661 (79%)	0.3883572 (2.500694e-01)	0.28837	1.055308e+07 (3.601370e+06)	1.380182e+01 (8.543694e-02)	1.147253e+00 (6.939737e-02)
ET7	2874.628	1,371,869 (33.5%)	0.3899542 (1.081437e-01)	0.37974	5.796267e+05 (8.409905e+04)	1.442739e+01 (7.504737e-02)	6.498523e-01 (6.543504e-02)

Table 10C. Using PWM initial guesses in MLE with Type 7 empirical quantile functions
In this case *p(k) = (k - 1) / (n - 1) = mode[F(x[k])]*.

Asset	Negloglik	Threshold (mode) Type=7	xi Estimator	xi initial guess (PWM7)	beta Estimator	logmu Estimator	logsd Estimator
ET2	3019.895	2,559,642 (20%)	0.2440805 (1.196974e-01)	0.241255	1.597878e+06 2.250094e+05)	1.520498e+01 (1.660845e-01)	5.000000e-01 (1.511915e-01)
ET4	3005.077	1,609,515 (70%)	0.7873854 (2.717887e-01)	0.616407	4.278366e+06 (1.314527e+06)	1.374089e+01 (8.232837e-02)	1.068265e+00 (6.992791e-02)
ET7	2873.552	1,375,373 (33.5%)	0.4572607 (1.151643e-01)	0.473569	4.901010e+05 (6.763896e+04)	1.443070e+01 (7.534806e-02)	6.519983e-01 (6.566100e-02)

Table 10D. Using PWM initial guesses in MLE with Type 5 empirical quantile functions
Here *p(k) = (k - 0.5) / n* is a **piecewise linear function** popular among
hydrologists where the knots are the values midway through the steps of the empirical cdf.

Asset	Negloglik	Threshold (Piecewise Linear: Type=5)	xi Estimator	xi initial guess (PWM5)	beta Estimator	logmu Estimator	logsd Estimator
ET2	3019.245	2,555,575 (20%)	0.2507767 (1.201940e-01)	0.246515	1.579561e+06 (2.228914e+05)	1.521016e+01 (1.622271e-01)	5.000000e-01
ET4	3005.195	1,614,552 (70%)	0.7893457 (2.727390e-01)	0.6170176	4.269634e+06 (1.317037e+06)	1.374201e+01 (8.242733e-02)	1.069579e+00 (7.001950e-02)
ET7	2874.000	1,448,153 (39%)	0.5168960 (1.289096e-01)	0.517495	4.551080e+05 (6.821365e+04)	1.437869e+01 (6.331044e-02)	6.120706e-01 (5.61599e-02)

does not hold for MOM initial guesses shown in tables 10A and 10B. Hence, it is not patently clear from the analysis whether using Type 7 or Type 5 empirical quantile function provides better results. For conservatism, be sure to specify which type your software is using.

For conservatism, we also recommend choosing an upper threshold of at least 0.975 (97.5%) based on an example from our current dataset with outliers removed. Table 11 shows that running an MLE analysis on ET3 with outliers removed and an upper threshold-percentile set at 95% seems to indicate a heavy-tailed distribution with infinite variance since the shape parameter is estimated at $\xi = 0.648$. However, if the threshold is raised to 97.5%, the initial guesses begin to drop precipitously as higher percentiles are considered. Table 12 shows that as estimations are made at higher thresholds, the initial guess for the shape parameter begins to plummet until the minimum negative loglikelihood is found with a shape-parameter estimate $\xi = 0.0$ implying that a lognormal distribution is perfectly adequate. Hence in Table 3, we have already shown ET3 as a loss event-type represented by a lognormal distribution. This fact would have been missed without setting a sufficiently high threshold.

Table 11: ET3 Parameters at Different Potential Maximum Thresholds (PW7)

Loss Type	Negloglik	Optimal Threshold Value	Maximum Upper Threshold Percentile	Initializer Type	xi (GPD shape)	beta (GPD scale)	logmu	logsd
ET3	2757.721	710,950 (65.5%)	0.95 (95%)	PWM7	0.64828 (0.1967)	3.219386e+05 (7.01975e+04)	13.1396 (0.06851)	0.870543 (0.05915)
ET3	2755.138	1,831,161 (97%)	0.975 (97.5%)	PWM7	0	3.857993e+06	13.1007	0.8117

Table 12. Contrasting ET3-Estimation with Maximum Threshold 95% vs 97.5%

percentile	negloglik	Initial- ξ Guess	Optimal- ξ Estimator 95% Maximum Threshold	Optimal- ξ Estimator 97.5% Maximum Threshold
0.650	2757.9714	0.602268		
0.655	2757.7206	0.609336	0.64827	
…				
0.940	2758.6183	0.64235		
0.945	2758.6751	0.60828		
0.950	2758.6247	0.56836		
Additional Rows from Increasing Threshold to 97.5% from 95%				
0.955	2758.4637	0.52068		
0.960	2757.8797	0.44046		
0.965	2756.5050	0.30652		
0.970	2755.4006	0.1161		0.0
0.975	2756.4860	-0.10242		

F. Fitting Copula Parameters: Step 2 under IFM

The second step in fitting a distributional copula under IFM is to estimate the copula parameters by loglikelihood (contingent upon the marginal parameters δ_i estimated in Step 1). The MLE method for estimating the copula parameters utilizes the copula density function. We will use the density function described in the subsection titled "Explicit Copulas: Determining the Density for the Gumbel Copula" to determine fitting for the Gumbel copula. Otherwise, we will use a hybrid version of IFM described in III.G to fit the parameters of standard Gaussian and t copulas by non-parametric methods.

There are two different approaches to deriving copula density functions depending upon whether the copula is implicit or explicit.

Implicit Copulas

For implicit copulas like the t-copula or Gaussian copula which lack explicit solutions, the chain rule of calculus may be applied directly to Sklar's theorem for the copula joint distribution. See equation (10) below.

Explicit Copulas

For explicit copulas (e.g. Archimedean copulas), we differentiate the copula function directly although the result may be complicated. See equations (18a)-(18d) below where we have derived the densities of the

Gumbel copula with dimensions 2 through 5. The resulting density functions serve as likelihood functions to determine (fit) the copula parameters. See equation (16) below for a bivariate example of the Gumbel copula's loglikelihood function.

Understanding the derivation of densities for implicit copulas greatly clarifies understanding meta-distributions (or meta-copulas) which were first discussed in III.A.1. Recollect that a meta-distribution (or meta-copula) combines the specified copula with *arbitrary* marginals. Therefore a software routine written in R, S-Plus, or Matlab to fit a *standard* copula may very well be inappropriate for fitting a meta-distribution (meta-copula). It is vital to understand the process of determining the density so the appropriate equations can be compared against prepackaged software routines for validity. In many cases, the analyst will need to write a modified software routine to capture correct results.

Implicit Copulas: Determining Density and Loglikelihood

Implicit standard copulas like the t-copula and the Gaussian copula are based on existing underlying distributions and do not have closed-form solutions. Because the marginals are chosen from the same family as the joint distributional family (e.g. the t-distribution or the Gaussian distribution), the standard copula may be written in terms of a simplification of Sklar's Theorem which states that a copula with marginal probabilities $u_1 = F(x_1)$ and $u_2 = F(x_2)$ equals the underlying joint distribution function $F(x_1, x_2)$:

(9a) $C(u_1, u_2) = C(F(x_1), F(x_2)) = F(x_1, x_2)$.

Alternatively, implicit *meta-copulas* may also be written using the generalized Sklar's Theorem where the marginals may be chosen *arbitrarily* as $u_1 = G_1(y_1)$ and $u_2 = G_2(y_2)$ from a family distinctly different from the joint distributional function $F(y_1, y_2)$:

(9b) $C(u_1, u_2) = C(G_1(y_1), G_2(y_2)) = F(y_1, y_2)$.

Note in (9b) that arbitrary marginals $u_i=G_i(y_i)$ have been substituted for the standard uniform random variables $u_i= F(x_i)$ usually serving as copula parameters due to the probability transform rule. For example, a joint distribution formed from a t-copula with arbitrary margins is known as a meta-t distribution which is characterized by a t-copula with arbitrary (e.g. lognormal) marginals. Similarly, a meta-Gaussian distribution is characterized by a Gaussian copula with arbitrary marginals.

To obtain the density $c(u_1, u_2)$ of an implicit bivariate meta-copula with *arbitrary* marginals (i.e. the density of a meta-copula), merely apply the chain rule from calculus to Sklar's equation in (9b)

$$\frac{\partial^2 C(u_1,u_2)}{\partial u_1 \partial u_2} = \frac{\partial^2 C(G_1(y_1),G_2(y_2))}{\partial u_1 \partial u_2}\left(\frac{\partial u_1}{\partial y_1}\right)\left(\frac{\partial u_2}{\partial y_2}\right) = \frac{\partial^2 F(y_1,y_2)}{\partial y_2 \partial y_1} \quad \Rightarrow$$

$$c(u_1,u_2) = c(G_1(y_1),G_2(y_2)) = \frac{f(y_1,y_2)}{g_1(y_1)g_2(y_2)} = \frac{f(G_1^{-1}(u_1),G_2^{-1}(u_2))}{g_1(G_1^{-1}(u_1))\, g_2(G_2^{-1}(u_2))} \Rightarrow \tag{10}$$

$$c(u_1,u_2) = \frac{f(G_1^{-1}(u_1),G_2^{-1}(u_2))}{g_1(G_1^{-1}(u_1))\, g_2(G_2^{-1}(u_2))}$$

since $\frac{\partial u_i}{\partial y} = \frac{\partial G_i(y)}{\partial y} = g_i(y)$ is the marginal density function associated with the ith arbitrarily selected marginal distribution function $G_i(y_i)$ and $f(y_1,y_2)$ is the joint density of the distribution function F.

Hence the implicit copula density (10) represents the ratio of the multivariate density function $f(y_1, y_2)$ derived from joint distribution function $F(y_1, y_2)$ implied by the copula family to the product of the selected (arbitrary) marginal densities $g_1(y_1)g_2(y_2)$ where all densities are evaluated using the quantile transforms of the selected marginals applied to the copula's standard uniform variable arguments.

Note that a *standard* implicit copula has margins from the same family as the copula itself. For example, the t-copula has a multivariate student-t distribution with student-t marginals. However, a meta-t copula is a t-copula with marginals from some family other than student-t distribution (like the lognormal or Gaussian distributions).

Recalling the basic loglikelihood function of the copula from (5), we see that the bivariate version (i=1,2) may be rewritten in terms of (10) as

$$\ln L(\theta;\hat{U}_1,...\hat{U}_n;\delta) = \sum_{t=1}^{n} \ln c_\theta(\hat{U}_t;\delta) = \tag{11}$$
$$\sum_{t=1}^{n} \ln f(G_1^{-1}(\hat{U}_{t,1};\delta_1), G_2^{-1}(\hat{U}_{t,2},\delta_2);\theta) - \sum_{t=1}^{n}\sum_{i=1}^{2} \ln g_i(G_i^{-1}(\hat{U}_{t,i};\delta_i);\delta,\theta)$$

where f( ) is the joint density associated with the copula family's joint distribution $F(X_1,X_2)$, G_i( ) are the arbitrary marginal distributions selected for the meta-copula model, g_i( ) are the marginal density functions associated with arbitrary marginal distributions G_i( ), and δ is the set of parameters associated with the marginal distributions estimated in step 1 of the IFM . It is important to distinguish that the second summation in (11) will be independent of the copula parameters θ if the marginal densities g_i( ) are not from the same family as the joint density f(). The differences in notation can be made clearer by examining a standard-t copula and a meta-t copula.

Implicit Copula Examples.

A standard t-copula example

Suppose we have a standard t-copula. Then the marginals are not arbitrary but are themselves univariate student-t distributions. Hence the t-copula combines a **joint-t density** and **univariate-t marginal densities** so the parameters are $\theta \equiv (\nu, P)$ where P is a d-dimensional correlation matrix among the d units of measure and ν is the degrees of freedom identical for each t-marginal. In this case the arbitrary marginal distributions G_i may be represented as standard-t marginals so the $\hat{U}_{t,i} = G_i(x_{t,i}) = t_\nu(x_{t,i})$ which implies that $t_\nu^{-1}(\hat{U}_{t,i}) = G_i^{-1}(\hat{U}_{t,i}) = x_{t,i}$. Similarly, the multivariate distribution may be written as $t_d(\nu,0,P)$ which implies the joint multivariate t density function is $g_{\nu,P}() = f(x_1,...x_d) = \dfrac{\partial^d(t_d(x_{i,t};\nu,0,P))}{\partial x_1...\partial x_d}$.

Finally write the density for the univariate student-t as $g_\nu() = \dfrac{\partial(t_\nu(x_{i,t}))}{\partial x_i}$. Substituting these values gives the loglikelihood function for a t-copula:

$$\ln L(\nu,P;\hat{U}_1,...\hat{U}_n) = \sum_{t=1}^{n} \ln g_{\nu,P}(t_\nu^{-1}(\hat{U}_{t,1}),...t_\nu^{-1}(\hat{U}_{t,d})) - \sum_{t=1}^{n}\sum_{i=1}^{d} \ln g_\nu(t_\nu^{-1}(\hat{U}_{t,i})). \tag{12}$$

Note that both sets of summands (across the joint density $g_{\nu,P}()$ and across the univariate densities $g_\nu()$ (all assumed identical)) impact the estimator of the degrees of freedom parameter ν since both the joint and marginal densities are distributed as student-t which is dependent on ν.

A meta-t with Gaussian marginals example:

In a meta-t distribution, we have a t-copula with marginals belonging to some family other than student-t. Assume we want Gaussian marginals. The loglikelihood equation for the meta-t copula will change significantly since the marginal distributions will no longer be student-t. In this case the arbitrary marginal distributions G_i may be represented as $\hat{U}_{t,i} = G_i(x_{t,i}) = \Phi(x_{t,i})$ which implies that $\Phi^{-1}(\hat{U}_{t,i}) = G_i^{-1}(\hat{U}_{t,i}) = x_{t,i}$ where Φ () is the univariate standard normal distribution function. In this case, the density of the univariate standard normal is $\phi() = \dfrac{\partial(\Phi(x_{i,t}))}{\partial x_i}$ so the loglikelihood equation for a meta-t distribution with IID Gaussian marginals may be written:

(13) $$\ln L(\nu, P; \hat{U}_1, ...\hat{U}_n) = \sum_{t=1}^{n} \ln g_{\nu,P}(\Phi^{-1}(\hat{U}_{t,1}), ..., \Phi^{-1}(\hat{U}_{t,d})) - \sum_{t=1}^{n}\sum_{i=1}^{d} \ln \phi(\Phi^{-1}(\hat{U}_{t,i})).$$

In (13), $g_{\nu,P}(\Phi^{-1}(\hat{U}_{t,1}), ..., \Phi^{-1}(\hat{U}_{t,d}))$ still ***denotes the* joint *density*** $f(G_1^{-1}(u_1), ..., G_d^{-1}(u_2))$ ***of a random vector with*** multivariate $t_d(\nu, 0, P)$ distribution where P is a linear correlation matrix and ν represents the degrees of freedom. However, in the meta-t case, the IID standard Gaussian marginal distributions replace the student-t marginal distributions and the Gaussian density replaces the student-t density. Hence the second summand in (13) (the double sum) does not depend upon ν, the degrees of freedom parameter.

Comparison of the student-t and meta-t copula cases

The first summands on the RHS of loglikelihood functions (12) and (13) are basically similar since each sums logarithms of a joint student-t distribution (although the joint distributions are evaluated at different arguments). However, a very significant difference occurs in the second term of the loglikelihood function.

When fitting a standard t-copula via (12), the second term of the loglikelihood function also influences the copula parameter estimation for ν, the degrees of freedom parameter, since the marginals are assumed to follow identical univariate-t distributions each with ν degrees of freedom. Contrarily, the second term in the meta-t distribution (13) has no influence on the ν-estimator since it contains only terms associated with Gaussian (or other arbitrary) marginals. Consequently, for analysts using the IFM method where the marginal parameters δ have been separately estimated in a first step, the general loglikelihood function in (11) may effectively be modified whenever fitting a meta-distribution by dropping the second summation.

Thus the same software functions cannot be employed to fit a standard t-copula and to fit a meta-t distribution (copula). The loglikelihood function for the t-copula contains summands across both multivariate and univariate-t densities which both depend upon ν, the degrees of freedom parameter, while only the first summand in the meta-t copula depends upon ν. It is absolutely vital not to call up canned computer routines to fit copulas without understanding what the underlying code does. This follows because most canned routines were built to handle straightforward copulas like the t-copula or Gaussian copula but not copulas for meta-distributions (like the meta-t which has a t-copula and arbitrary marginals).

Shortcuts to MLE in Implicit Copulas

Some implicit copulas allow shortcuts to the application of MLE. In other words, an easier mechanism can be substituted. For example the Gauss Copula $C_P^{Ga}(u)$ has a single parameter for calibration, the correlation matrix containing the linear (Pearson's) correlation coefficients on the off-diagonals. It has been proved that a good estimator of the linear correlation matrix parameter is the Spearman's rho matrix (untransformed).

The t-Copula $C_{\nu,P}^{t}(u)$ has multiple parameters, ν (the degrees of freedom), and P (the correlation matrix) which must be calibrated to data. For the t-copula, a ***transformation*** of the Kendall's tau matrix may provide a good estimator of the linear correlation matrix parameter. Each element r_{ij} of the correlation matrix P may be estimated as $r_{ij} = \sin[\pi \rho_\tau(x_i, x_j)/2]$ where $\rho_\tau(x_i, x_j)$ is Kendall's tau for rank-correlations.

Explicit Copulas: Determining the Density for the Gumbel Copula

The Gumbel copula (sometimes referred to as the Gumbel-Hougaard copula) is a function of a single parameter θ and has a closed-form solution. It exhibits positive upper tail dependency and hence is a conservative candidate for operational-risk copulas. Its major drawback is the single parameter describing all units of measure (event loss types). Perhaps due to the complicated formula, existing software programs generally contain code to evaluate only

the two-dimensional Gumbel copula.[29] Expansion to five or seven dimensions is desirable to cover all event types in operational risk. In the following equations, we derive the density function for the Gumbel copula in higher dimensions. The resulting Gumbel copula densities may be readily coded in a statistical software language like R, S-Plus, or Matlab to serve as the basis for loglikelihood functions needed to fit the copula to data. We have coded an algorithm for the Gumbel loglikelihood function in the R-language to identify the parameter θ and fit the Gumbel copula. Results from fitting the Gumbel copula to actual data appear in Part III, section I.

The n-dimensional Gumbel copula may be defined as:

(14a) $y = C(u_i, ... u_j) = \exp\{-[z_j(u)]^{1/\theta}\}$

where

(14b) $z_j(u) = \sum_{i=1}^{j} (-\ln u_i)^{\theta}, \quad d_i = \frac{(-\ln u_i)^{\theta-1}}{u_i}, \quad \frac{\partial z}{\partial u_i} = (-\theta/u_i)(-\ln u_i)^{\theta-1} = -\theta\, d_i.$

Differentiation of the copula function yields the copula density $c(u_1, ..., u_n)$. Consider the two-dimensional case where n=2.

$$\frac{dy}{du_i} = d_i\, y\, [z_j(u)]^{\frac{1}{\theta}-1} \Rightarrow$$

$$\frac{\partial^2 y}{\partial u_1 \partial u_2} = d_1 y \frac{\partial([z_2(u)]^{\frac{1}{\theta}-1})}{\partial u_2} + [z_2(u)]^{\frac{1}{\theta}-1} \frac{\partial y}{\partial u_2} \Rightarrow$$

$$c(u_1, u_2) = d_1[y(\frac{1}{\theta}-1)([z_2(u)]^{\frac{1}{\theta}-2})(-\theta\, d_2) + d_2 y([z_2(u)]^{\frac{2}{\theta}-2})]$$

Therefore the density for the two-dimensional Gumbel copula may be written as:

(15) $c(u_1, u_2) = d_1 d_2\, y\, [z_2(u)]^{\frac{1}{\theta}-2} \{[z_2(u)]^{\frac{1}{\theta}} + (\theta - 1)\}.$

The corresponding loglikelihood function sums the log of the density across all n observations in each of the pairs (u1, u2):

(16)
$$\ln L(\theta; u_{1,t}, u_{2,t}) = \sum_{t=1}^{n} \ln c(u_{1,t}, u_{2,t}) =$$
$$\sum_{t=1}^{n}\sum_{i=1}^{2} \ln(d_i) + \ln(y_t) + (\frac{1}{\theta} - 2)\ln[z_2(u_t)] + \ln\{[z_2(u_t)]^{\frac{1}{\theta}} + (\theta - 1)\}$$

Define the constants a_i as in (17).

(17) $a_0 = \theta - 1, a_1 = 2\theta - 1, a_2 = 3\theta - 1, ... a_j = (j+1)\theta - 1.$

Applying similar analysis to higher dimensional copulas provides appropriate representations for Gumbel copula density functions for dimensions 2-5.

Two-dimensional Gumbel copula density:

[29] E.g. the QRMlib software accompanying the text by McNeil, Frey, Embrechts (2005) can fit only a two-dimensional Gumbel copula since that is the maximum density available. Another library by Aas and Frigessi is also limited to two dimensions. A new R-language library called Gumbel now has a routine gumbel::dgumbel() which can do three-dimensional densities.

(18a) $c(u_1, u_2) = \quad d_1 d_2 y\,[z_2(u)]^{\frac{1}{\theta}-2}\{[z_2(u)]^{\frac{1}{\theta}} + a_0\}$

Three-dimensional Gumbel copula density:

(18b) $c(u_1, u_{2,} u_3) = d_1 d_2 d_3 y\,[z_3(u)]^{\frac{1}{\theta}-3}\{[z_3(u)]^{\frac{2}{\theta}} + 3a_0[z_3(u)]^{\frac{1}{\theta}} + a_1 a_0\}$

Four-dimensional Gumbel copula density:

(18c)
$$c(u_1, u_{2,} u_{3,} u_4) = d_1 d_2 d_3 d_4 y\,[z_4(u)]^{\frac{1}{\theta}-4}\{[z_4(u)]^{\frac{3}{\theta}} + 6a_0[z_4(u)]^{\frac{2}{\theta}} + a_0[3(a_2 - 1) + a_1][z_4(u)]^{\frac{1}{\theta}} + a_2 a_1 a_0\}$$

Five-dimensional Gumbel copula density:

(18d)
$$c(u_1, u_{2,} u_{3,} u_4, u_5) = d_1 d_2 d_3 d_4 d_5 y\,[z_5(u)]^{\frac{1}{\theta}-5}\{[z_5(u)]^{\frac{4}{\theta}} + 10a_0[z_5(u)]^{\frac{3}{\theta}} + a_0[6(a_3 - 2) + 3(a_2 - 1) + a_1][z_5(u)]^{\frac{2}{\theta}} + a_0\{a_1 a_2 + (a_3 - 1)[3(a_2 - 1) + a_1]\} + a_3 a_2 a_1 a_0\}$$

G. Fitting Copula Parameters: a Hybrid IFM Step 2 for Simulated Losses

This paper proposes to empirically measure diversification benefits to capital by simulating 100,000 joint annual loss-observations where dependency characteristics are described alternatively by a Gaussian, student-t, or Gumbel copula and the marginal distributions for loss event types are lognormal-GPD mixtures. Empirical quantiles of the simulated annual losses for all event types will then be used to infer the diversification ratio in equation (2b) along with VaR and other important statistics.

Currently available software routines for copula simulations generate realizations ***U*** of marginal probabilities (not losses) from *standard* (rather than meta) copulas. ***U*** is known as a matrix of pseudo-sample probabilities. Existing algorithms generally proceed in a two-step methodology and require parameters of the standard copula as inputs[30]:

- simulate a multivariate set of losses ($X_{i,t}$) from the joint distribution function associated with the *standard* copula;
- determine the marginal probabilities $u_{i\,t}$ associated with the generated losses by applying the *standard* copula's marginal distribution functions as $u_{1t} = F(X_{1t})$ *and* $u_{2t} = F(X_{2t})$

Hence current simulation algorithms require knowledge of parameters for the standard copula rather than the meta-copula.

As can be readily seen by comparing Sklar's Theorem for standard-copulas (equation 9a) versus the theorem for meta-copulas (equation 9b), losses from a meta-copula can be generated by applying the quantile transforms of a meta-copula's arbitrary margins to the probabilities u_i simulated from a standard copula.[31] For a bivariate copula,

[30] For specific descriptions of simulations from standard Gaussian and t copulas, see Algorithms 5.9 and 5.10 (p. 193) in McNeil, Frey, Embrechts (2005). For a description of simulation from a Gumbel copula, see Algorithm 5.48 on p. 224 of the same text..

[31] Nelsen's text (2006) provides an argument for the quantile transform in Exercise 2.15, p. 29 which states the following. Let X_1 and Y_1 be random variables with continuous distribution functions F_1 and G_1, respectively, and copula C. Let F_2 and G_2 be

note that for every loss pair ($X_{1,t}$, $X_{2,t}$) generated by the first step in a standard copula simulation algorithm and every corresponding probability pair (u_{1t}, u_{2t}) generated in the second step of the algorithm, a corresponding pair of outcomes (Y_{1t}, Y_{2t}) could be generated from the meta-distribution's arbitrary marginals (G_{1t}, G_{2t}) to create the same set of marginals probabilities:

$$u_{1t} = G_1(Y_{1t}) = F(X_{1t}) \quad and \quad u_{2t} = G_2(Y_{2t}) = F(X_{2t}) .$$

Hence to determine the losses (Y_{1t}, Y_{2t}) from a meta-copula associated with pseudo-probabilities generated by an associated standard copula, merely apply the appropriate quantile transforms of the meta-marginals to each of the simulated standard copula probabilities (u_{1t},u_{2t}):

$$\begin{aligned} G_1^{-1}(u_{1t}) &= G_1^{-1}(F(X_{1t})) = Y_{1t} \\ G_2^{-1}(u_{2t}) &= G_2^{-1}(F(X_{2t})) = Y_{2t} \end{aligned} .$$

Hence, a vector of simulated losses for each time period is produced by applying appropriate quantile transforms (using parameters for the marginals previously estimated in step 1 of the IFM) to the simulated probabilities:

$$\mathbf{Y} = \{G_1^{-1}(\mathbf{u}_1), G_2^{-1}(\mathbf{u}_2)\} .$$

McNeil, Frey, Embrechts (2005) provide a description of the preceding argument in their QRM textbook (p. 193):

> "Assume that the problem of generating realizations ***U*** from a particular copula has been solved. The converse of Sklar's Theorem shows us how we can sample from interesting meta distributions that combine this copula with an arbitrary choice of marginal distributions. If ***U*** has df C, then we use quantile transformation to obtain $X \equiv (F_1^{\leftarrow}(U_1),...,F_d^{\leftarrow}(U_d))'$ which is a random vector with margins F_1,…F_d and multivariate df $C(F_1(x_1),...,F_d(x_d))$. This technique is extremely useful in Monte Carlo studies of risk...".

Thus a procedure to simulate losses from a meta-copula requires fitting sample loss data to a standard copula rather than a meta-copula in the second-step of a hybrid IFM. The probabilities from the simulated standard copula for all units of measure are then transformed to simulated losses by applying the quantile transforms for the associated marginals whose parameters were determined in Step 1 of the IFM (section III.E.)

Hence in this hybrid step, we choose to fit parameters to a standard copula by using a non-parametric approach where the pseudo-sample probabilities input to the fitting function are identified from the empirical distribution function.

Assume we are trying to fit a standard t-copula which requires estimating parameters of a copula with identical student-t marginals for each unit of measure (ET). One way to emulate this case using sample data is with non-parametric estimation, merely applying the empirical distribution function (edf) to each sample loss vector $X_{i,t}$ representing event-type i to form a pseudo-sample matrix of probabilities. Each edf vector within the resulting matrix contains a set of identical probabilities but with different ordering in the respective column. Hence the non-parametric edf-matrix emulates drawings from a set of identical t-marginals as required. Thus passing the non-parametric pseudo-sample to a standard-t copula fitting routine which solves equation (12) for the degrees of freedom and correlation parameters should provide appropriate standard t-copula estimators.[32]

Simulated losses from the corresponding meta-copula may then be obtained from the simulated probabilities of the standard copula by applying the appropriate quantile transform for each marginal distribution as previously described in this section. For the lognormal-GPD mixture, the quantile function will be the inverse of equation (7) if the first stage of IFM identified the marginal as "heavy-tailed"; otherwise it will be the inverse of a regular

---

another pair of continuous distribution functions and set $X_2 = F_2^{-1}(F_1(X_1))$ *and* $Y_2 = G_2^{-1}(G_1(Y_1))$. Then (a) the distribution functions of X_2 and Y_2 are F_2 and G_2 respectively; and (b) the copula of X_2 and Y_2 is C.

[32] Creating a pseudo-sample by applying t-marginals to observed losses from a lognormal cannot work because the one-sided lognormal loss distribution can reflect losses only on the right-hand side of a t-distribution.

lognormal. The parameters associated with each quantile function and estimated in the first IFM step have already been recorded in Table 2 (for the full-data case) and in Table 3 (for the case with outliers removed).

H. Simulating the Losses Via a Distributional Copula Function

Section III.E described the separate estimation of marginal distribution parameters for each unit of measure (ET) under the first step of the IFM. Section III.G described fitting the proposed copula parameters to probability-data in the second-step of a hybrid IFM method contingent upon the empirical distribution function of the observed losses. Given knowledge of both sets of fitted parameters, we can simulate sequences of 52 weekly losses from the proposed copula for each ET which can then be summed as an annual loss. We repeat the simulation 100,000 times using the following steps.

1) **Select a standard-copula and corresponding marginal distribution functions for all ETs**.
   We propose two sets of separate generator models:
   - non-conservative models with straight lognormal marginals for each ET
     - meta-Gaussian copula
     - meta-t copula
     - meta-Gumbel copula
   - conservative models with lognormal-GPD mixtures (heavy tails) for each ET
     - meta-Gaussian copula
     - meta-t copula
     - meta-Gumbel copula

   Note we must create meta-distributions even in the lognormal case since operational-loss distributions have one-sided support and hence are inconsistent with the two-sided student-t or Gaussian marginals associated with the standard copulas.

   The first set of lognormal (non-conservative) generator models should serve as a radical lower limit on the dependency structure since lognormality implies none of the ETs is assumed to have a heavy tail. As pointed out by Giacometti et al (2008), such a model will overestimate the degree of diversification provided by dependency since it lacks any extreme losses from heavy-tailed distributions. Hence "choice of the loss distribution thus becomes of central concern and must be estimated with a high degree of accuracy".

   The second set of copulas will provide conservative estimates of the dependency reduction when compared to the medium-tailed lognormal marginals. The t- and Gumbel-copulas provide additional conservatism since they reflect positive tail dependency (absent from the Gaussian copula) which may potentially increase the required capital by reflecting large simultaneous losses in pairs of loss event types. Each ET will reflect a separate set of parameter estimators for its corresponding marginal distribution. The number of ETs for any bank will depend on the size of the individual bank's data and whether synchronicity of weekly data allowed sufficient observations for estimation in each ET.
2) Fit each column of historical weekly loss data (associated with a separate ET) against the lognormal-gpd mixture in equation (8) using MLE, identifying five parameters for each ET. If the estimated threshold is too high to achieve convergence, re-estimate the ET as a straight lognormal distribution.
3) Fit the historical weekly loss data to the chosen standard-copula by non-parametrically calculating the pseudo-sample probability matrix using the empirical distribution function and passing the edf to an MLE/optimization fitting routine as an argument. This is like a hybrid IFM step 2. Review the discussion in III.G. This step generates the parameters for a standard-copula rather than a meta-copula.
4) **Simulate 100,000 years of *weekly* loss data for each copula generator model**.
   - For each year, simulate a matrix of 52 weeks of losses from each loss event type, creating a matrix with 52 rows and up to 7 columns of event loss types depending upon whether sufficient historical data was available to estimate parameters for the marginal loss distributions. The loss simulation consists of two steps previously described in Part III.G.
     - First simulate a marginal loss probability matrix from a standard copula (not a meta-copula) using copula parameters previously fitted in step 3. Estimated parameters for the various copulas are presented in Tables 14A and 14B in Appendix 3.
     - Secondly, apply a quantile function from the appropriate lognormal-GPD mixture (or straight lognormal) loss distribution to each column of probabilities in the matrix from the previous

step to create simulated losses. The parameters associated with the marginal's loss distribution must have been previously estimated by applying MLE to the weekly historical loss data in step 2. This step creates a matrix of simulated weekly losses for 52 weeks. Estimated parameters appropriate for quantile functions of the mixture distributions are contained in Table 2 for the full dataset and in Table 3 for the dataset with outliers removed. Parameters for straight lognormal cases are contained in Tables 13A and 13B in Appendix 3.

- **Aggregate the simulated weekly losses** for each ET in a year to get a single annual loss in each ET for the 52 weeks in year
- **Aggregate the sum of the annual losses across event types to get an aggregate annual bank loss each year**

5) **Calculate the empirical quantiles** {0.95, 0.99, 0.999} for **aggregate** annual bank losses from the 100,000 year sample (this is the term *totalOpVaR* in equation (2b))
6) **Calculate empirical quantiles** {0.95, 0.99, 0.999} for **each ET**'s annual simulated loss in each year of the 100,000 year sample and sum the ET quantiles (this is the term *subadditivity upper bound* in equation (2b))
7) **Divide** the subadditivity upper bound from step 4 by the totalOpVaR from step 3 and subtract 1 to get the diversification ratio defined in equation (2b).

I. Empirical Diversification Results

Table 15.
Diversification Ratios, OpVaR, and Expected Shortfall
Comparing Heavy- and Thin-Tailed Marginals
Comparing Full Data Set to Data with Outliers Removed

Loss Distri-bution & Initializer Method	Copula Type	$VaR_{99}(\sum_{i=1}^{5} ET_i)$ VaR at 99th percentile	$\sum_{i=1}^{5} VaR_{99}(ET_i)$ Subadditivity upper bound at 99th percentile	Diversi-fication Ratio[33] (99th per-centile)	$VaR_{999}(\sum_{i=1}^{5} ET_i)$ VaR at 99.9th percentile	$\sum_{i=1}^{5} VaR_{999}(ET_i)$ Subadditivity upper bound at 99.9th percentile	Diversi-fication Ratio (99.9th per-centile)	$ES_{99.9}$ Expected shortfall at 99.9th percentile	$pES_{99.9}$ Subadditive upper bound for expected shortfall
All LN Fit	Full Data ($Billions)								
	Gauss	0.2088835	0.2265731	8.4687%	0.334323	0.3570227	6.7896%	0.44038	0.46987
	T	0.2116417	0.2292060	8.2991%	0.340682	0.3665826	7.6026%	0.44377	0.47532
	Gumbel	0.2093754	0.2246699	7.3048%	0.342613	0.3639409	6.2252%	0.49341	0.51976
	Outliers Removed ($Billions)								
	Gauss	0.1362306	0.1504573	10.4431%	0.1751789	0.1955252	11.615%	0.20129	0.22599
	T	0.1358662	0.1500447	10.4356%	0.1749272	0.1970870	7.6026%	0.20717	0.23176
	Gumbel	0.2076021	0.2255529	8.6467%	0.3315575	0.3589521	8.2624%	0.47723	0.50725
4 Mixtures + 1 LN LN-GPD Mixtures PWM7 Initializers 3 Mixtures + 2 LN	Full Data ($Billions)								
	Gauss	6.516076	6.601120	1.305%	84.78458	85.3146	0.6252%	483.99	485.48
	T	6.558567	6.612461	0.8217%	82.92779	83.51742	0.7110%	2693.3	2695.5
	Gumbel	6.494234	6.509054	0.2282%	107.02168	107.59066	0.5317%	5985.4	5987.18
	Outliers Removed ($Billions)								
	Gauss	0.465092	0.498738	7.2342%	1.942176	2.108336	8.5553%	1.2286	1.3193
	T	0.457959	0.488307	6.6227%	2.096548	2.236126	6.6574%	4.8944	5.3158
	Gumbel	0.462863	0.492723	6.4513%	2.268555	2.372386	4.5770%	13.4611	13.8741

Empirical results for fitting the marginal distributions of ET1, ET2, ET3, ET4, ET7 as lognormal-GPD mixtures for a single bank have already been shown in Tables 2 and 3 (see Part III.E).

[33] This column represents the diversification ratio from equation (2b) multiplied by 100 to give the diversification benefit as a percentage at the 99th percentile. The third column further to the right represents the diversification ratio as a percentage for the 99.9th percentile.

Empirical results for fitting the marginal distributions of ET1, ET2, ET3, ET4, ET7 as straight lognormal distributions appear as Tables 13A and 13B in Appendix 3.

The results of fitting copulas to pseudo-sample probabilities generated by edf representations of weekly loss data for the single bank appear as Tables 14A and 14B in Appendix 3.

Table 15 presents the results of the diversification ratios (capital reductions) at the 99th and 99.9th percentiles simulated from Gauss, t, and Gumbel distributional copulas (using the method described in III.H) for the full dataset and for the dataset with four outliers removed from ET4. When using all data points, four of the five loss event types are well-fitted by lognormal-GPD mixtures. When four outliers are removed from the dataset, only three of the five loss event types are still well-fitted by the heavy-tailed mixtures.

The table also presents results under the extremely non-conservative assumption that marginal distributions for all loss event types are lognormally distributed (i.e. are medium-tailed distributions) even though the more conservative heavy-tailed assumption is empirically supported by the data.

It also contrasts results using a fitted Gaussian copula with no tail dependence versus more conservative meta-t and Gumbel copulas which may exhibit positive (upper) tail dependence increasing the likelihood of simultaneous large losses in pairs of ETs and hence reducing "diversification".

We draw the following conclusions from data for this single bank from results in Tables 14 and 15.

- The VaR results show no subadditivity violations (not even at lower percentiles like the 95th which is not shown in the table).
- The levels of required capital without removing outliers seems prohibitively high when some (four of five) ETs have "heavy-tailed" distributions.
- The diversification (capital-reduction) ratios of 7%-11% are substantially higher for all copula types when every marginal is artificially fit to a "medium-tailed" (lognormal) distribution.
- The diversification (capital-reduction) ratios are 4.5%-8.5% with outliers removed and three of five marginal distributions reflecting "heavy-tails". This range is substantially lower than the 20-25% found for a single European bank by Giacometti et al (2008) (and the 30-50% reductions found in previous studies using thin-tailed distributions by other authors like Dalla Valle (2008) and Chapelle (2004)). It is clearly important to investigate data from additional banks.
- When using the sample with outliers removed, the diversification ratios produced by the t-copula and Gumbel-copula are substantially lower than that provided by the Gauss-copula
- The Gumbel copula seems to fit the data well and provides consistent results even though it is a single-parameter model.
- The Gumbel copula results in the highest capital requirement in all cases at the 99.9th percentile.
- The Gumbel copula results in substantially higher Expected Shortfalls for all cases where some marginals have "heavy tails".
- The positive upper tail dependence property of the Gumbel copula seems to have its strongest effects at very high quantiles like the 99.9th where it induces a higher capital requirement. The removal of outliers apparently dropped the Gumbel's estimated positive upper tail dependency coefficient from 0.09129 to 0.07169 which may partially explain why the 99.9% VaR from the Gumbel copula is proportionally much larger than the VaR from the t-copula with all data included relative to when outliers are removed. The decline in tail dependency may reflect the simultaneous removal of four large values (largest, 21st largest, 23rd largest, and 128th largest of 196 observations) from ET2 when the four ET4 outliers were removed.

J. Summary

This paper has presented strong graphical evidence that one-sided support for operational-risk losses eliminates one of the primary sources of diversification, negative tail dependency, from large simultaneous losses. Hence operational-risk losses should exhibit substantially lower diversification benefits than returns with two-sided support in market risk. Consequently, analysts should continue to practice conservatism required by the U.S. Rule

until the factors driving diversification in operational-risk losses can be more thoroughly understood. Any scenario analysis cannot be based on logic derived from market- or credit-risk where underlying data has two-sided support.

This paper attempts to measure the diversification (capital-reduction) ratio implied by historical LDCE2004 loss data from a single large U.S. bank. To measure the diversification effect, the paper attempts to simulate annual losses using a distributional copula and use the empirical distribution function of the simulated data to measure VaR and other related statistics.

Simulation from a distributional copula requires identification of parameter for both the copula and corresponding marginal distributions. This paper utilizes a hybrid of the two-stage Inference Functions for Marginals (IFM approach) popularized by Joe for parameter identification. Using aggregated weekly loss data to avoid synchronicity problems with sparse data, the first stage estimates marginal distribution parameters for historical (LDCE2004) loss data using MLE under the assumption of lognormal-GPD mixture distributions for loss data which reflect "heavy tails". The second stage of the hybrid approach applies non-parametric estimation to the sample losses using the empirical distribution function to build a pseudo-sample matrix of probabilities which emulate drawings from identical marginals required to fit a standard copula. Using the edf matrix as input, we fit standard Gaussian, t, and Gumbel copulas via MLE. To fit the Gumbel copula, we derived mathematical formulations (see equations 18a-18d) for the Gumbel copula density and corresponding likelihood function for dimensions up to five (the number of ET with sufficient data) and coded a corresponding optimization algorithm since existing computer code provided densities of maximum dimension two or three. A sample of 100,000 annual losses was constructed by simulating and summing consecutive sets of 52 weekly losses. Each weekly loss was calculated by simulating a set of probabilities (52 x 5) from the previously fitted standard (Gaussian, t, or Gumbel) copula and then applying quantile transforms for the lognormal-GPD mixtures to the simulated probabilities in each column of the matrix. Annual losses in each ET and across the firm are then readily attainable by summation. The diversification (capital reduction) ratio and other VaR measures can be easily extracted from the empirical distribution function of the 100,000 simulated annual losses.

The simulation results suggest diversification ratios of only 4.5%-8.5% from all copula models at the 99.9th percentile which is substantially lower than results found in other studies.

Future Extensions

Distributional Copula Approaches with larger numbers of parameters than the Gumbel:

- Pair copula cascade (Bedford and Cooke) Model selection may be critical and challenging (Aas)
- Nested Archimedean Copulas (McNeil) Requires constraints on copula parameters

Testing data from many additional banks to determine if the capital reduction is consistent or is highly affected by the historical loss data.

Testing banks with sufficient data for all seven event-loss types.

Developing a Bayesian strategy

Appendix 1. Tail dependency

Joe (1997) initially defined two types of asymptotic **positive** tail dependence indicating whether a pair of random variables exhibit ***similar* extreme (tail) behavior**. Hence positive tail dependence implies that a bivariate pair of extreme tail results are either both very large or very small. Tail dependence measures do not depend on the underlying marginals and hence can be written in terms of the joint distribution's underlying distributional copula.

- Positive *upper* tail dependence is measured as

$$\lambda_u^+ = \lim_{q \uparrow 1^-} P\,(X_2 > F_2^{-1}(q) \mid X_1 > F_1^{-1}(q)) = \lim_{q \uparrow 1^-} \frac{\hat{C}(1-q, 1-q)}{1-q} = \lim_{q \uparrow 1^-} \frac{1-2q+C(q,q)}{1-q}.$$

  It measures the propensity for X_2 to exceed a very high quantile (fall into the upper tail) given that X_1 also exceeds the same high quantile (upper tail) value. Hence λ_u^+ measures joint outcomes in the ***upper right corner*** of a graph of bivariate outcomes. It is applicable to random variables defined with either two-sided

infinite support $X \in \Re = (-\infty, \infty)$ as in market or credit-risk and variables with single-sided infinite support $X \in \Re = (0, \infty)$ like operational-risk losses since simultaneous large values are not precluded by either type of support.

- Positive *lower* tail dependence is measured as:

$$\lambda_l^+ = \lim_{q \downarrow 0} P(X_2 \le F_2^{-1}(q) \mid X_1 \le F_1^{-1}(q)) = \lim_{q \downarrow 0} \frac{C(q,q)}{q}.$$

  It measures the propensity for X_2 to fall below a very small quantile in the lower tail given that X_1 lies below the same small quantile. Hence λ_l^+ measures joint outcomes in the ***lower left corner*** of a graph of bivariate outcomes. It is basically irrelevant for diversification in random variables with single-sided infinite support $X \in \Re = (0, \infty)$ like operational-risk losses since it describes only joint outcomes which are both close to zero.

Zhang (2008) introduced measures of **negative** (asymptotic) tail dependence indicating whether a pair of random variables exhibits ***opposite* extreme (tail) behavior**. The existence of negative tail dependence is vital to understanding diversification of portfolio stock returns with two-sided infinite support whereas it is relatively unimportant in describing diversification of operational risk losses which have one-sided positive support.

- ***Negative upper*** tail dependence is measured as

$$\lambda_u^- = \lim_{q \uparrow 1^-} P\,(X_2 > F_2^{-1}(q) \mid X_1 \le F_1^{-1}(1-q)) = 1 - \lim_{q \uparrow 1^-} \frac{C(1-q,\, q)}{1-q}.$$

  It measures the propensity for X_2 to exceed a very high quantile (fall into the upper tail) given that X_1 falls below the corresponding (1-q) percentile in the opposite tail. Hence λ_u^- measures joint outcomes in the *upper left corner* of a graph of bivariate outcomes. However, for variables with one-sided support, the upper left corner displays a pair of outcomes with a large positive value and a value close to zero and hence no diversification benefit..
- ***Negative lower*** tail dependence is measured as:

$$\lambda_l^- = \lim_{q \downarrow 0} P(X_2 \le F_2^{-1}(q) \mid X_1 > F_1^{-1}(1-q)) = 1 - \lim_{q \downarrow 0} \frac{C(1-q,q)}{q}.$$

  It measures the propensity for X_2 to fall below a very small quantile in the lower tail given that X_1 lies above the corresponding large (1-q) quantile in the upper tail. Hence λ_l^- measures joint outcomes in the *lower right corner* of a graph of bivariate outcomes. However, for variables with one-sided support, the lower right corner displays a pair of outcomes with a large positive value and a value close to zero and hence no diversification benefit.

Operational-risk loss distributions exhibit purely positive support $X_i \in \Re = [0, \infty)$ rather than two-sided support. The one-sided support drastically limits the impact of negative (or opposite) movements in the variable pair. With two-sided support, a large positive value can offset a large negative value, creating “diversification” effects.

Furthermore, strict Archimedean copulas with continuous marginals have no negative tail dependence (i.e. $\lambda_u^- = 0 = \lambda_u^+$). Hence Gumbel and Clayton copulas are appropriate for modeling marginals with one-sided support like operational-risk loss distributions. Alternatively, copulas associated with elliptical distributions (e.g. the t-copula with t-marginals) have positive values for negative tail dependence and hence are pertinent for modeling marginals with two-sided support where large losses may be offset by large gains, providing "diversification".

The Gauss copula precludes tail dependencies. The Gumbel copula has positive upper tail dependency in a bivariate setting which implies a positive conditional probability that if one risk factor exceeds a large quantile, the second risk factor also exceeds the same quantile value. Hence upper tail dependency represents the opposite of diversification.
"The essential determinant or whether the copula of a normal variance mixture has tail dependence or not is the tail of the distribution of the mixing variable W….If W has a distribution with a power tail, then we get tail dependence; otherwise we get asymptotic independence….a consequence of a general result for elliptical distributions given in section 7.3.3." QRM, p. 212

See Example 5.33 (eqn. 5.31) which shows that the t-distribution's **positive** tail dependence measure can be stated as

$$\lambda = 2t_{\nu+1}(-\sqrt{\frac{(\nu+1)(1-\rho)}{(1+\rho)}})$$

where ρ is the linear correlation coefficient. ρ may be replaced by the estimate of Kendall's tau to give a simple measure of the tail dependence for the t-copula

$$\rho_{12} = \sin(\frac{\pi}{2}\rho_\tau)$$

where ρ_τ is estimated from data in the R-language by the function Kendall() in QRMlib or the function cor() in the stats package to obtain Kendall's ρ_τ = tauK = cor(data.t.2d, method="kendall") where ρ_{12} will be an off-diagonal element of the resulting tauK matrix

Appendix 2

Levy Copula Model

In the Levy model, individual aggregate loss distributions for each unit of measure and the sum of aggregate losses across all units of measure must be compound Poisson processes.[34] Furthermore, the Levy measure $\Pi_i([0,x)) = \lambda_i F_i(x_i)$ for each operational-risk loss cell i is totally determined by the expected Poisson frequency of loss occurrences and the severity distribution of the cell (unit of measure). Note the Levy measure has total mass equal to the expected number of loss arrivals λ since as $x \to \infty$, $F_i(x) \to 1$ so $\Pi_i([0,\infty)) = \lambda_i$. The Levy **marginal tail integral** is the product of the ith cell's Poisson frequency and the tail of the ith severity distribution: $\overline{\Pi}_i([x,\infty)) = \lambda_i \overline{F}_i(x_i)$ and is therefore a transform of the Levy measure. Consequently, the Levy copula (for compound Poisson processes with finite positive jumps) defines the **joint tail integral** of the d-dimensional compound-Poisson aggregate loss process across the d units of measure. Specifically, the Levy copula separates the marginal tail integrals and dependencies in the jump measures from the joint tail integral:
$\overline{\Pi}(x_1,...x_d) = C^{Levy}(\overline{\Pi}_1(x_1),,,\overline{\Pi}_d(x_d))$.
Since the Levy copula maps tail integrals into the non-negative real line (i.e. it maps $(0,\infty]^d \to [0,\infty]$), it is not a distributional copula.[35]

The primary importance of the Levy copula model is its amenability to investigating dependence between frequency and severity at the same time because the Levy measures Π and corresponding tail integrals $\overline{\Pi}$ are independent of time and depend upon the frequency of losses only through its expected value. Hence modeling of dependence between coincident losses in different units of measure (e.g. losses in the same week in different cells) can be readily accomplished.

Pareto-Levy Copula Model

[34] This behavior implies the multivariate LDA model is invariant to the structure of the units of measure.
[35] Instead, it might be described as a tail-integral copula. Note the Levy copula's domain is sometimes described as the punctured positive cone because it excludes 0 as a loss value.

The Pareto-Levy copula modifies the Levy copula model by transforming the tail integral to be the tail function of a standard Pareto distribution: $\overline{\Pi}_i([x_i,\infty)) \equiv \overline{\Gamma}_i([x_i,\infty)) = \overline{F}_i^{Pareto}(x_i) = x_i^{-1},\ x > 0.$ Hence the tail integral depends only upon a continuous tail (severity) distribution. This transformation introduces an infinite number of jumps (infinite variation) over the time period, meaning the summed aggregate loss distribution is no longer compound-Poisson. Hence the Pareto-Levy copula is specified as

$\overline{\Gamma}([x_1,\infty),...[x_d,\infty)) = C^{L-P}(\frac{1}{\Gamma_1(x_1)},...\frac{1}{\Gamma_1(x_1)}) = C^{L-P}(x_1,,,x_d)$. The Pareto-Levy copula is also not a distributional copula since it maps $(0,\infty]^d \rightarrow [0,\infty]$. It may be described as a tail-integral copula where the tail integral is precisely the tail function of a standard Pareto distribution. The Pareto-Levy copula separates the joint Levy tail measure into a dependency structure and marginal Levy tail measures.

Appendix 3: Fitting Results

The results for fitting straight lognormals to the full dataset appear in Table 13A:

Table 13A

Loss Type	Negloglik	logmu	logsd
ET1	2739.905	12.47165 (0.078039)	1.0925560 (0.055182)
ET2	3135.761	15.14338 (0.040656)	0.5691939 (0.02874)
ET3	2834.664	13.14586 (0.064487)	0.9028297 (0.045599)
ET4	3134.278	14.07424 (0.117535)	1.6454956 (0.08311)
ET7	2963.164	14.26903 (0.040403)	0.5656492 (0.028569)

Results for fitting straight lognormals to the dataset with outliers removed appears in Table 13B.

Table 13B

Loss Type	Negloglik	logmu	logsd
ET1	2670.743	12.45337 (0.074949)	1.038532 (0.05299)
ET2	3030.369	15.11470 (0.034074)	0.472146 (0.024094)
ET3	2763.510	13.116233 (0.062621)	0.867714 (0.044280)
ET4	3018.616	13.95119 (0.102599)	1.42166 (0.072548)
ET7	2895.868	14.259808 (0.039761)	0.550956 (0.028116)

Results for fitting the lognormal-GPD mixtures to the full dataset were shown in Table 2 in the main body of the paper.

RResults for fitting the lognormal-GPD mixtures to the dataset with outliers removed were shown in Table 3 in the main body of the paper.

Table 14A. Fitting Standard Copulas to Full Data Using Non-parametric EDFs for Pseudo-sample

Full data, LN-GPD mixture			
copula	**negloglik**	**nu (df)**	**Correlation matrix**
t-copula	-28.54851	8.027427	$P [,1] [,2] [,3] [,4] [,5] [1,] 1.00000000 0.07647740 0.1672030 0.03528907 0.04773063 [2,] 0.07647740 1.00000000 0.2030429 0.19457791 0.02837321 [3,] 0.16720300 0.20304290 1.0000000 0.17757638 0.19195029 [4,] 0.03528907 0.19457791 0.1775764 1.00000000 0.16252182 [5,] 0.04773063 0.02837321 0.1919503 0.16252182 1.00000000
Gaussian-copula	-23.15775	NA	$P [,1] [,2] [,3] [,4] [,5] [1,] 1.00000000 0.08978377 0.2188185 0.06109352 0.09640038 [2,] 0.08978377 1.00000000 0.2104894 0.23468559 0.07661365 [3,] 0.21881848 0.21048945 1.0000000 0.21098863 0.22380747 [4,] 0.06109352 0.23468559 0.2109886 1.00000000 0.19622207 [5,] 0.09640038 0.07661365 0.2238075 0.19622207 1.00000000
copula	**negloglik**	**theta**	**positive upper tail dependency** λ_u^+
Gumbel copula	-11.91119	1.072273 (0.02334683)	0.09129

Table 14B. Fitting Standard Copulas to Data with Outliers Removed Using Non-parametric EDFs

Outliers Removed, LN-GPD mixture			
copula	**negloglik**	**nu (df)**	**Correlation matrix**
t-copula	-19.62375	11.51919	$P [,1] [,2] [,3] [,4] [,5] [1,] 1.00000000 0.06197525 0.1700540 0.02900597 0.05739806 [2,] 0.06197525 1.00000000 0.1743880 0.16839145 0.01895638 [3,] 0.17005405 0.17438799 1.0000000 0.13747018 0.17014580 [4,] 0.02900597 0.16839145 0.1374702 1.00000000 0.16064263 [5,] 0.05739806 0.01895638 0.1701458 0.16064263 1.00000000
Gaussian-copula	-18.21555	NA	$P [,1] [,2] [,3] [,4] [,5] [1,] 1.00000000 0.06886389 0.2184339 0.07032646 0.15314437 [2,] 0.06886389 1.00000000 0.1268797 0.17762663 0.03093256 [3,] 0.21843386 0.12687973 1.0000000 0.15322317 0.17606315 [4,] 0.07032646 0.17762663 0.1532232 1.00000000 0.18756119 [5,] 0.15314437 0.03093256 0.1760632 0.18756119 1.00000000
copula	**negloglik**	**theta**	**positive upper tail dependency** λ_u^+
Gumbel copula	-3.770828	1.055587 (0.02392866)	0.07169

Bibliography

Bee, “On Maximum Likelihood Estimation of Operational Loss Distributions”, University of Trento Discussion Paper, 2005.

Bee, “Copula-based multivariate models with applications to risk management and insurance”, University of Trento Discussion Paper, 2005.

Cech, "Copula-based top-down approaches in financial risk aggregation", working paper, December 2006.

Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer Verlag 2001.

Dalla Valle, Fantazzini, Giudici, “Copulae and Operational Risks”, *International Journal of Risk Assessment and Management*, 2008.

De Fontnouvelle, Rosengren, Jordan, 2004, "Implications of Alternative Operational Risk Modeling Techniques", Federal Reserve Bank of Boston.

Dupuis, “Exceedances over High Thresholds: a Guide to Threshold Selection”, *Extremes*, Vol. 1, No. 3, 1999, pp. 251-261.

Dutta, K. and Perry, J. (2006), “A tale of tails: an empirical analysis of loss distribution models for estimating operational risk capital”, Federal Reserve Bank of Boston, Working Paper No 06-13.

El Gamal, Inanoglu, Stengel, "Multivariate Estimation for Operational Risk with Judicious Use of Extreme Value Theory", *Journal of Operational Risk*, Vol. 2, Number 1, 2007.

Embrechts, Lambrigger, and Wuthrich, "Multivariate Extremes and the Aggregation of Dependent Risks: Examples and Counter-Examples", *Extremes*, 2008.

Embrechts, Neslehova, Wuthrich, “Additivity Properties for Value-at_Risk under Archimedean Dependence and Heavy-Tailedness”, *Insurance: Mathematics and Economics*, 2008.

Genest and Favre, “Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask”, *Journal of Hydrologic Engineering*, Vol. 12, No. 4, July 1, 2007.

Giacometti, Rachev, Chernobai, Bertocchi, "Aggregation Issues in Operational Risk", *Journal of Operational Risk*, Vol. 3, Number 3, 2008.

Hosking and Wallace, "Parameter and Quantile Estimation for the Generalized Pareto Distribution", *Technometrics*, August 1987, Vol. 29, No. 3.

Joe, H. (1997) *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.

Juarez and Schucany, "Robust and Efficient Estimation for the Generalized Pareto Distribution", *Extremes*, Vol. 7, No. 3, September 2004.

McNeil, Frey, Embrechts, *Quantitative Risk Management*, Princeton Press, 2005.

Moscadelli, 2003, "Modeling of Operational Risk: Experience with Analysis of Data Collected by the Basel Committee".

Nelsen (2006), *An Introduction to Copulas*, 2nd Edition, Springer.

Nešlehová, Embrechts, and Chavez-Demoulin, (2006), “Infinite mean models and the LDA for operational risk”, *Journal of Operational Risk* 1(1), 3–25.

Rosenberg and Schuermann, “A General Approach to Integrated Risk Management with Skewed, Fat-tailed Risks”, Working Paper, 2005.

Zhang, "Modeling total tail dependence along diagonals", *Insurance: Mathematics and Economics* 42, 73–80, 2008

Xu, J. J., "Statistical Modeling and Inference for Multivariate and Longitudinal Discrete Response Data", 1996 unpublished Ph.D. thesis, Department of Statistics, University of British Columbia.