Global Derivatives & Risk Management 2001

Innovation Approaches For Mastering Correlation, Copulas & Multivariate Models And Optimising The Application To Credit Risk, Market Risk And Risk Integration

Led by Alexander McNeil, ETH and
Professor Ruediger Frey,
University of Maryland

Seminar: Friday 11th May 2001
Hotel Les Ambassadeur, Juan les Pins

Organised by:
ICBI
www.icbi-uk.com

Correlation, Copulas, Multivariate Models with Applications to Market Risk, Credit Risk and Risk Integration

Prof. Alexander McNeil, Department of Mathematics, ETH Zurich
Prof. Rüdiger Frey, Swiss Banking Institute, University of Zurich

Global Derivatives & Ri$k Management 2001

Juans-Les-Pins, South of France, 11th May 2001

mcneil@math.ethz.ch, freyr@isb.unizh.ch

http://www.math.ethz.ch/~mcneil
http://www.math.ethz.ch/~frey

I. Understanding the Fundamentals of Modelling Dependent Risks

- Exploring the basics of multivariate statistics
- A detailed assessment of multivariate normal distributions
- Evaluating elliptical models and normal mixture models
- Examining portfolio theory in an elliptical world

©2001 (Alexander McNeil and Rüdiger Frey)

1. Multivariate Statistics: Basics

Some Univariate Notation

Let X be a random variable (rv) representing a risk or risk factor. Let F be the distribution function (df) of X, i.e. $F(x) = P(X \leq x)$.

The *tail* or *survivor function* is denoted $\overline{F}(x) = P(X > x)$.

Where it exists the density of X is written $f(x)$ and satisfies $F(x) = \int_{-\infty}^{x} f(u)du$.

For $q \in (0,1)$ the qth quantile of the distribution of X is denoted $F^{-1}(q)$. This should be taken to mean the generalised quantile $F^{-1}(q) = \inf\{x \in \mathbb{R} : F(x) \geq q\}$.

Where appropriate we may interpret this quantile as a *Value-at-Risk* of the risk X and write $\text{VaR}_q(X) = F^{-1}(q)$.

Multivariate Notation

Let $\mathbf{X} = (X_1, \ldots, X_d)'$ be a d-dimensional *random vector* representing risks of various kinds. Possible interpretations:

- returns on d financial instruments (market risk)
- asset value returns for d companies (credit risk)
- results for d lines of business (risk integration)

An individual risk X_i has *marginal* df $F_i(x) = P(X_i \leq x)$.
A random vector of risks has *joint* df

$$F(\mathbf{x}) = F(x_1, \ldots, x_d) = P(X_1 \leq x_1, \ldots, X_d \leq x_d)$$

or joint survivor function

$$\overline{F}(\mathbf{x}) = \overline{F}(x_1, \ldots, x_d) = P(X_1 > x_1, \ldots, X_d > x_d).$$

Multivariate Models

If we fix F (or $\overline{F}$) we specify a *multivariate model* and implicitly describe both the marginal behaviour and the *dependence structure* of the risks.

Calculating Marginal Distributions

$$F_i(x_i) = P(X_i \leq x_i) = F(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty),$$

i.e. limit as arguments tend to infinity.

In a similar way *higher dimensional marginal* distributions can be calculated for other subsets of $\{X_1, \ldots, X_d\}$.

Independence

$X_1, \ldots, X_d$ are said to be mutually independent if

$$F(\mathbf{x}) = \prod_{i=1}^{d} F_i(x_i), \quad \forall \mathbf{x} \in \mathbb{R}^d.$$

©2001 (Alexander McNeil and Rüdiger Frey)

Densities of Multivariate Distributions

Most, but not all, of the models we consider can also be described by joint densities $f(\mathbf{x}) = f(x_1, \ldots, x_d)$, which are related to the joint df by

$$F(x_1, \ldots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(u_1, \ldots, u_d) du_1 \ldots du_d.$$

Existence of a joint density implies existence of marginal densities $f_1, \ldots, f_d$ (but not vice versa).

Equivalent Condition for Independence

$$f(\mathbf{x}) = \prod_{i=1}^{d} f_i(x_i), \quad \forall \mathbf{x} \in \mathbb{R}^d$$

©2001 (Alexander McNeil and Rüdiger Frey)

2. The Multivariate Normal (or Gaussian) Distribution

This distribution can be defined by its density

$$f(\mathbf{x}) = (2\pi)^{-d/2}|\Sigma|^{-1/2} \exp\left\{-\frac{(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}\right\},$$

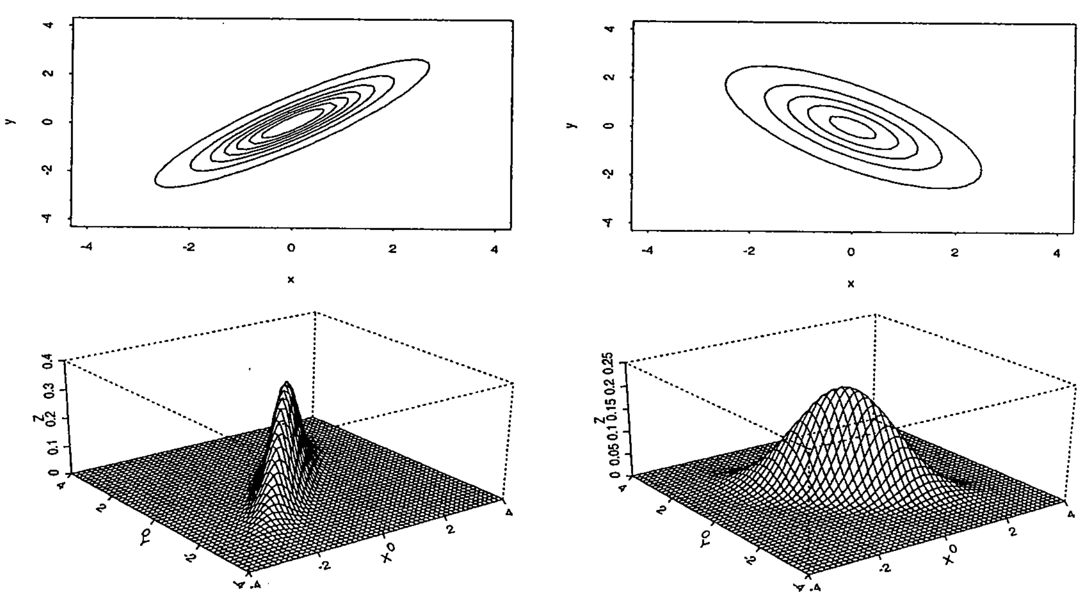
where $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ is a positive definite matrix.

- If $\mathbf{X}$ has density f then $E(\mathbf{X}) = \boldsymbol{\mu}$ and $\text{cov}(\mathbf{X}) = \Sigma$, so that $\boldsymbol{\mu}$ and Σ are the *mean vector* and *covariance matrix* respectively. A standard notation is $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$.

- Clearly, the components of $\mathbf{X}$ are mutually independent if and only if Σ is diagonal. For example, $\mathbf{X} \sim N_d(\mathbf{0}, \mathbf{I})$ if and only if $X_1, \ldots, X_d$ are *iid* $N(0, 1)$.

©2001 (Alexander McNeil and Rüdiger Frey)

Bivariate Standard Normals ($\rho = 0.9$, $\rho = -0.7$)



©2001 (Alexander McNeil and Rüdiger Frey)

Properties of Multivariate Normal Distribution

- The marginal distributions are univariate normal.
- Linear combinations $\mathbf{a}'\mathbf{X} = a_1 X_1 + \cdots a_d X_d$ are univariate normal with distribution $\mathbf{a}'\mathbf{X} \sim N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\Sigma\mathbf{a})$.
- Conditional distributions are multivariate normal.
- The sum of squares $(\mathbf{X} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \chi_d^2$ (chi-squared).

Limitations of Multivariate Normal Distribution

- Tails are very thin - few extreme events.
- Simultaneous extremes in several margins relatively infrequent.
- Very strong symmetry (known as elliptical symmetry).

3. Multivariate Normal Mixture Distributions

Multivariate Normal Variance-Mixtures

Let $\mathbf{Z} \sim N_d(\mathbf{0}, \Sigma)$ and let W be an *independent*, positive, scalar random variable. Let $\boldsymbol{\mu}$ be any deterministic vector of constants. The vector $\mathbf{X}$ given by

$$\mathbf{X} = \boldsymbol{\mu} + W\mathbf{Z}$$

is said to have a multivariate normal variance-mixture distribution.

Easy calculations give $E(\mathbf{X}) = \boldsymbol{\mu}$ and $\mathrm{cov}(\mathbf{X}) = E(W^2)\Sigma$.
Correlation matrices of $\mathbf{X}$ and $\mathbf{Z}$ are identical: $\mathrm{corr}(\mathbf{X}) = \mathrm{corr}(\mathbf{Z})$.

Multivariate normal variance mixtures provide the most useful examples of so-called *elliptical* distributions.

Examples of Multivariate Normal Variance-Mixtures

Multivariate t with ν degrees of freedom. Take $W = \sqrt{\nu/V}$, $V \sim \chi^2_\nu$.

Symmetric generalised hyperbolic. Let W have a NIG distribution.

Multivariate Normal Mean-Variance-Mixtures

We can generalise the construction of the previous slide as follows:

$$\mathbf{X} = g(W) + W\mathbf{Z}, \quad g : \mathbb{R}^+ \to \mathbb{R}^d$$

This gives us a larger class of distributions, but in general they are no longer elliptical and corr$(\mathbf{X}) \neq$ corr$(\mathbf{Z})$.

Example: full class of generalised hyperbolic distributions.
$g(W) = \boldsymbol{\mu} + W^2\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is deterministic vector and W is NIG.

The Multivariate t Distribution

This distribution can also be defined by its density

$$f(\mathbf{x}) = k_{\Sigma,\nu,d}\left(1 + \frac{(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}{\nu}\right)^{-\frac{(\nu+d)}{2}}$$

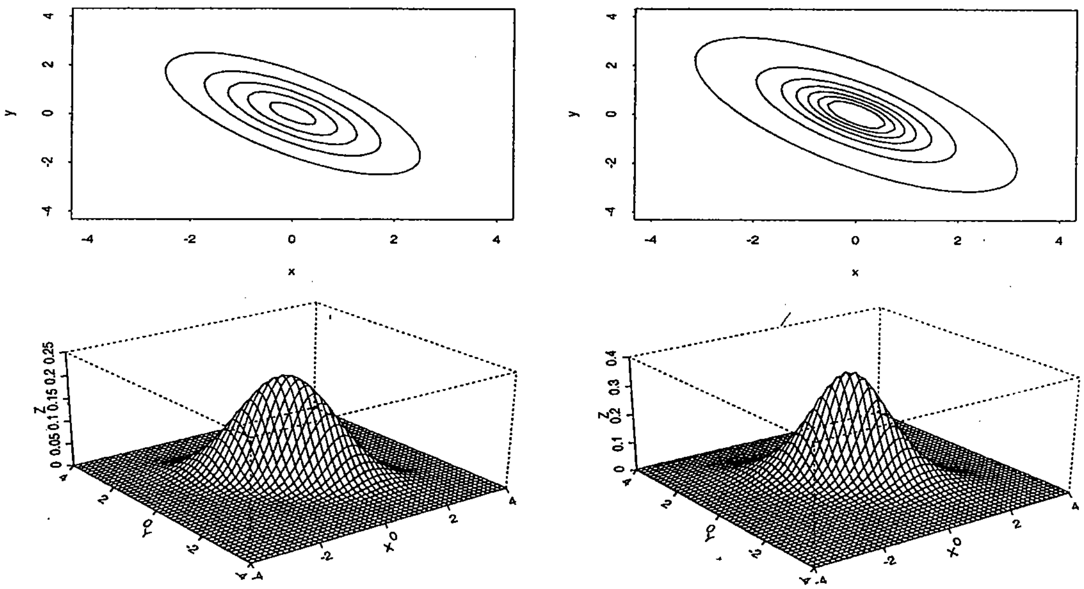
where $\boldsymbol{\mu} \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d\times d}$ is a positive definite matrix,ν is the degrees of freedom and $k_{\Sigma,\nu,d}$ is a normalizing constant.

- If $\mathbf{X}$ has density f then $E(\mathbf{X}) = \boldsymbol{\mu}$ and cov$(\mathbf{X}) = \frac{\nu}{\nu-2}\Sigma$, so that $\boldsymbol{\mu}$ and Σ are the *mean vector* and *dispersion matrix* respectively. For finite variances (and defined correlations) $\nu > 2$. Our notation is $\mathbf{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$.

- If Σ is diagonal the components of $\mathbf{X}$ are *uncorrelated*. They are not independent.

- The multivariate t distribution has heavy tails.

Bivariate Normal and t ($\rho = -0.7$, $\nu = 3$, variances equal 1)



Elliptical distributions

A random vector $(Y_1, \ldots, Y_d)$ is *spherical* if its distribution is invariant under rotations, i.e. for all $U \in \mathbb{R}^{d \times d}$ with $U'U = UU' = I_d$

$$\mathbf{Y} \stackrel{\mathrm{d}}{=} U\mathbf{Y}.$$

A random vector $(X_1, \ldots, X_d)$ is called elliptical if it is an affine transform of a spherical random vector $(Y_1, \ldots, Y_d)$,

$$\mathbf{X} = A\mathbf{Y} + \mathbf{b},$$

$A \in \mathbb{R}^{d \times d}$, $\mathbf{b} \in \mathbb{R}^d$.

General Remark:
If $\mathbf{X}$ has covariance matrix Σ, then $\mathrm{cov}(A\mathbf{X}) = A\Sigma A'$.

Properties of Elliptical Distributions

- The density of an elliptical distribution is constant on ellipsoids.
- Many of the nice properties of the multivariate normal are preserved. In particular, all linear combinations $a_1X_1 + \ldots + a_dX_d$ are of the same type.
- All marginal distributions are of the same type.
- Linear correlation matrices successfully summarise dependence, since mean vector, covariance matrix and the distribution type of the marginals determine the joint distribution uniquely.

Elliptical Distributions and Risk Management

Consider set of linear portfolios of elliptical risks
$\mathcal{P} = \{Z = \sum_{i=1}^{d} \lambda_i X_i \mid \sum_{i=1}^{d} \lambda_i = 1\}$.

1. VaR is a *coherent* risk measure in this world. It is monotonic, positive homogeneous (P1), translation preserving (P2) and, most importantly, *sub-additive*

   $\mathrm{VaR}_\alpha(Z_1 + Z_2) \leq \mathrm{VaR}_\alpha(Z_1) + \mathrm{VaR}_\alpha(Z_2)$, for $Z_1, Z_2 \in \mathcal{P}, \alpha > 0.5$.

2. Among all portfolios with the same expected return, the portfolio minimizing VaR, or any other risk measure ϱ satisfying
   P1 $\varrho(\lambda Z) = \lambda\varrho(Z),\ \lambda \geq 0$,
   P2 $\varrho(Z + a) = \varrho(Z) + a,\ a \in \mathbb{R}$,
   is the Markowitz variance minimizing portfolio.
   Risk of portfolio takes the form $\varrho(Z) = E(Z) + \text{const} \cdot \mathrm{sd}(Z)$.

4. Other Multivariate Distributions and Concepts

Other Continuous Multivariate Distributions

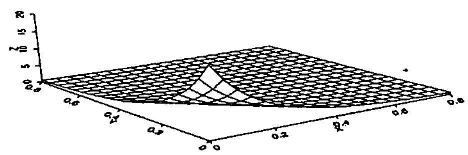
There are infinitely many non-elliptical distributions and in such models covariance and correlation are less natural concepts.

Example: multivariate Pareto distribution

$\mathbf{X}$ has a multivariate Pareto if its survivor function is given by

$$\overline{F}(x_1, \ldots, x_d) = \left(\sum_{i=1}^{d} \frac{x_i + \beta_i}{\beta_i} - d + 1 \right)^{-\alpha}, \quad \mathbf{x} > 0, \alpha > 0, \boldsymbol{\beta} > 0,$$

where α is the *tail index* and $\beta_1, \ldots, \beta_d$ are scaling parameters.
This kind of *skewed, heavy-tailed* distribution may be appropriate for positive-valued loss data.



©2001 (Alexander McNeil and Rüdiger Frey)

Multivariate Discrete Distributions

There are many of these including multivariate binomials, Bernoullis and Poissons.

Example: Bivariate Poisson (common shock model)
Let N_1 and N_2 be numbers of credit losses in two portfolios over 1 year horizon.
Suppose there are caused by three kinds of event:
Event 1 causes losses of type 1
Event 2 causes losses of type 2
Event 3 causes losses of both types 1 and 2
Assume: events occur as independent Poisson processes with yearly rates λ_1, λ_2 and λ_3.

$$N_1 \sim \text{Po}(\lambda_1 + \lambda_3) \quad N_2 \sim \text{Po}(\lambda_2 + \lambda_3)$$

(N_1, N_2) has a bivariate Poisson distribution.
$N_1 + N_2$ is *not* Poisson (but distribution can be calculated by convolution or Panjer recursion).
Distribution described by *probability function* $P(N_1 = n_1, N_2 = n_2)$.

©2001 (Alexander McNeil and Rüdiger Frey)

Factor Models

In a factor model we attempt to reduce the effective dimensionality of $\mathbf{X}$. $\mathbf{X}$ is said to follow a p-factor model with $p < d$ if

$$X_i = \sum_{j=1}^{p} a_{i,j}\Theta_j + \epsilon_i, \quad i = 1, \ldots, d,$$

or, in vector form, $\mathbf{X} = A\Theta + \epsilon$.

Notation:
$\Theta = (\Theta_1, \ldots, \Theta_p)'$ is a random vector of factors;
$\epsilon = (\epsilon_1, \ldots, \epsilon_d)'$ is a random vector of uncorrelated terms;
$a_{i,j}$ and A represent deterministic factor loadings;
$(\Theta_1, \ldots, \Theta_p)$ is uncorrelated with $(\epsilon_1, \ldots, \epsilon_d)$.

Typical Applications:
$\mathbf{X}$ are stock returns; Θ are index returns.
$\mathbf{X}$ are firm asset values; Θ are country and industry effects.

Factor Models II

Note that assumption of factor model is equivalent to saying covariance matrix Σ of $\mathbf{X}$ has the form $\Sigma = \Lambda\Lambda' + \Psi$, for some $d \times p$ dimensional matrix Λ and some diagonal matrix Ψ.

Statistical Techniques for Estimating Factor Models

Multivariate Regression Models
If Θ are known measurable factors then we could also consider our factor model to be a multivariate regression model. By collecting repeated observations of $\mathbf{X}$ and Θ we could estimate A.

Factor Analysis
If Θ represent abstract factors, in whose existence we believe but whose identities are unclear to us, we can use the technique of factor analysis. We collect repeated observations of $\mathbf{X}$ and try to estimate Λ and Ψ in decomposition above.

References

On general multivariate statistics:

- Mardia, Kent & Bibby (1979). Multivariate Analysis. Academic Press, London.
- Seber (1984). Multivariate Observations. Wiley, New York.

On multivariate distributions:

- Kotz, Balakrishnan, Johnson (2000). Continuous Multivariate Distributions. Wiley, New York.
- Johnson, Kotz, Balakrishnan (1997). Discrete Multivariate Distributions. Wiley, New York.

II. Advanced Dependence Concepts: Copulas and Extremal Dependence

- Describing dependence with copulas
- Understanding the limitations of correlation
- An assessment of alternative dependence measures
- Tail dependence and dependent extreme values
- Survey of useful copula families

1. Describing Dependence with Copulas

On Uniform Distributions

Lemma 1 (*probability transform*)
Let X be a random variable with *continuous* distribution function F. Then $F(X) \sim U(0,1)$ (standard uniform).

$$P(F(X) \leq u) = P(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u, \quad \forall u \in (0,1).$$

Lemma 2 (*quantile transform*)
Let U be uniform and F the distribution function of *any* rv X. Then $F^{-1}(U) \stackrel{\text{d}}{=} X$ so that $P(F^{-1}(U) \leq x) = F(x)$.

These facts are the key to all statistical simulation and essential in dealing with copulas.

©2001 (Alexander McNeil and Rüdiger Frey)

A Definition

A copula is a *multivariate uniform distribution*, or the df thereof.

Notation: $C : [0,1]^d \to [0,1]$

Properties

- Uniform Margins
  $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all $i \in \{1, \ldots, d\}$, $u_i \in [0,1]$

- Fréchet Bounds

$$\max\left\{\sum_{i=1}^{d} u_i + 1 - d, 0\right\} \leq C(\mathbf{u}) \leq \min\{u_1, \ldots, u_d\}.$$

Remark: right hand side is df of $\overbrace{(U, \ldots, U)}^{\text{d times}}$, where $U \sim U(0,1)$.

©2001 (Alexander McNeil and Rüdiger Frey)

Sklar's Theorem (general statement)

Let F be a joint distribution function with margins $F_1, \ldots, F_d$. There exists a copula such that for all $x_1, \ldots, x_d$ in $[-\infty, \infty]$

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)).$$

If the margins are continuous then C is unique; otherwise C is uniquely determined on $\text{Ran}F_1 \times \text{Ran}F_2 \ldots \times \text{Ran}F_d$.

Proof (case of continuous margins)

$$\begin{aligned} F(x_1, \ldots, x_d) &= P(X_1 \leq x_1, \ldots, X_d \leq x_d) \\ &= P(F_1(X_1) \leq F_1(x_1), \ldots, F_d(X_d) \leq F_d(x_d)) \\ &= C(F_1(x_1), \ldots, F_d(x_d)). \end{aligned}$$

Henceforth, unless explicitly stated, vectors $\mathbf{X}$ will be assumed to have continuous marginal distributions.

Copulas and Dependence Structures

Sklar's theorem shows how a unique copula C fully describes the dependence of $\mathbf{X}$. This motivates a further definition.

Definition: Copula of $\mathbf{X}$
The copula of $(X_1, \ldots, X_d)$ (or F) is the df C of $(F_1(X_1), \ldots, F_d(X_d))$.

We sometimes refer to C as the dependence structure of F.

Invariance
C is invariant under strictly increasing transformations of the marginals. If $T_1, \ldots, T_d$ are strictly increasing, then $(T_1(X_1), \ldots, T_d(X_d))$ has the same copula as $(X_1, \ldots, X_d)$.

Examples of copulas

- *Independence.*
  $X_1, \ldots, X_d$ are mutually independent $\iff$ their copula C satisfies $C(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i$.

- *Comonotonicity - perfect dependence.*
  $X_i \stackrel{\text{a.s.}}{=} T_i(X_1)$, T_i strictly increasing, $i = 2, \ldots, d$, $\iff$ C satisfies $C(u_1, \ldots, u_d) = \min\{u_1, \ldots, u_d\}$.

- *Countermonotonicity - perfect negative dependence.* (d=2)
  $X_2 \stackrel{\text{a.s.}}{=} T(X_1)$, T strictly decreasing, $\iff$ C satisfies $C(u_1, u_2) = \max\{u_1 + u_2 - 1, 0\}$.

Copulas Implicit in Well-Known Distributions

In every multivariate df with continuous marginals there is a unique implicit copula given by

$$C(u_1, \ldots, u_d) = F\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right).$$

Gaussian Copula: $\mathbf{X}$ standard multivariate normal with correlation matrix R.

$$\begin{aligned} C_R^{\text{Ga}}(u_1, \ldots, u_d) &= P\left(\Phi(X_1) \le u_1, \ldots, \Phi(X_d) \le u_d\right) \\ &= P\left(X_1 \le \Phi^{-1}(u_1), \ldots, X_d \le \Phi^{-1}(u_d)\right) \end{aligned}$$

where Φ is df of standard normal. $R = I$ gives independence; as $R \to J$ we get comonotonicity.

In the same way we can extract copulas from multivariate normal mixture distributions, to get for example *t copulas* ($C_{\nu,R}^{\text{t}}$) or *generalised hyperbolic copulas*.

Archimedean Copulas ($d = 2$)

- Gumbel Copula

$$C_\beta^{Gu}(u_1, u_2) = \exp\left(-\left((-\log u_1)^\beta + (-\log u_2)^\beta\right)^{1/\beta}\right).$$

$\beta \geq 1$: $\beta = 1$ gives independence; $\beta \to \infty$ gives comonotonicity.

- Clayton Copula

$$C_\beta^{Cl}(u_1, u_2) = \left(u_1^{-\beta} + u_2^{-\beta} - 1\right)^{-1/\beta}.$$

$\beta > 0$: $\beta \to 0$ gives independence ; $\beta \to \infty$ gives comonotonicity.

- Frank Copula

$$C_\beta^{Fr}(u_1, u_2) = -\frac{1}{\beta}\log\left(1 + \frac{(e^{-\beta u_1} - 1)(e^{-\beta u_2} - 1)}{e^{-\beta} - 1}\right).$$

$\beta \neq 0$: $\beta \to -\infty$ gives countermonotonicity; $\beta \to 0$ gives independence; $\beta \to \infty$ gives comonotonicity.

©2001 (Alexander McNeil and Rüdiger Frey)

Archimedean Copulas - Construction and Extensions

All our Archimedean copulas have the form

$$C(u_1, u_2) = \psi^{-1}(\psi(u_1) + \psi(u_2)),$$

where $\psi : [0, 1] \mapsto [0, \infty]$ is strictly decreasing and convex with $\psi(1) = 0$ and $\lim_{t \to 0} \psi(t) = \infty$.

The simplest higher dimensional extension is

$$C(u_1, \ldots, u_d) = \psi^{-1}(\psi(u_1) + \cdots + \psi(u_d)).$$

Example: Gumbel copula: $\psi(t) = -(\log(t))^\beta$

$$C_\beta^{\mathrm{Gu}}(u_1, \ldots, u_d) = \exp\left(-\left((-\log u_1)^\beta + \cdots + (-\log u_d)^\beta\right)^{1/\beta}\right).$$

These copulas are *exchangeable* (invariant under permutations). Other extensions with more parameters possible, but complex.

©2001 (Alexander McNeil and Rüdiger Frey)

Useful Copula Families

- *Elliptical Copulas*: rich in parameters - parameter for every pair of variables; easy to simulate.
- *Archimedean Copulas*: closed forms - appealing for calculations; but not rich in parameters.
- *Extreme Value Copulas*: arise naturally in multivariate extreme value theory; satisfy $C^t(u_1, \ldots, u_d) = C(u_1^t, \ldots, u_d^t)$, $\forall t > 0$, Example: Gumbel.
- *Extremal Copulas*: are copulas of vectors whose components are all either pairwise comonotonic or countermontonic; rank correlation matrix consists of 1's and -1's.
  Let J be subset of $\{1, \ldots, d\}$. General form of extremal copula:

$$C(u_1, \ldots, u_d) = \max\left\{\min_{i \in J} u_i + \min_{j \in J^c} u_j - 1, 0\right\}.$$

  Example: comonotonicity copula.

2. Understanding Limitations of Correlation

Drawbacks of Linear Correlation

Denote the linear correlation of two random variables X_1 and X_2 by $\rho(X_1, X_2)$. We should be aware of the following.

- Linear correlation only gives a scalar summary of (linear) dependence and $\text{var}(X_1), \text{var}(X_2)$ must exist.
- X_1, X_2 independent $\Rightarrow \rho(X, Y) = 0$.
  But $\rho(X_1, X_2) = 0 \not\Rightarrow X_1, X_2$ independent.
  Example: spherical bivariate t-distribution with ν d.f.
- Linear correlation is not invariant with respect to strictly increasing transformations T of X_1, X_2, i.e. generally

$$\rho(T(X_1), T(X_2)) \neq \rho(X_1, X_2).$$

Fallacies in the Use of Correlation

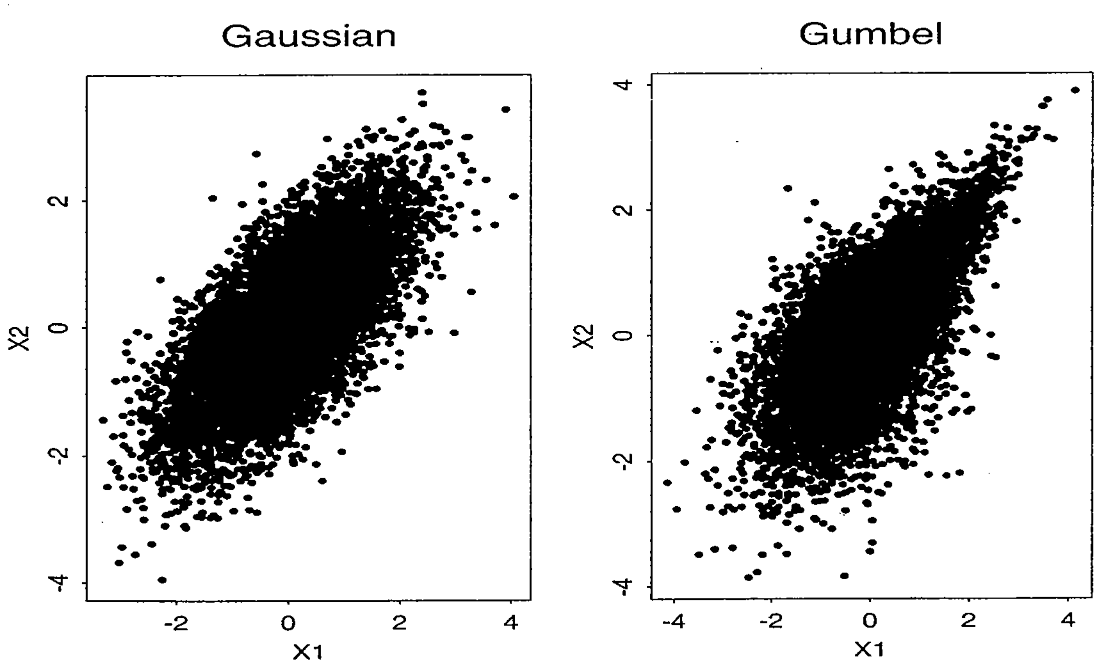
We consider random vectors $(X_1, X_2)'$.

Fallacy 1

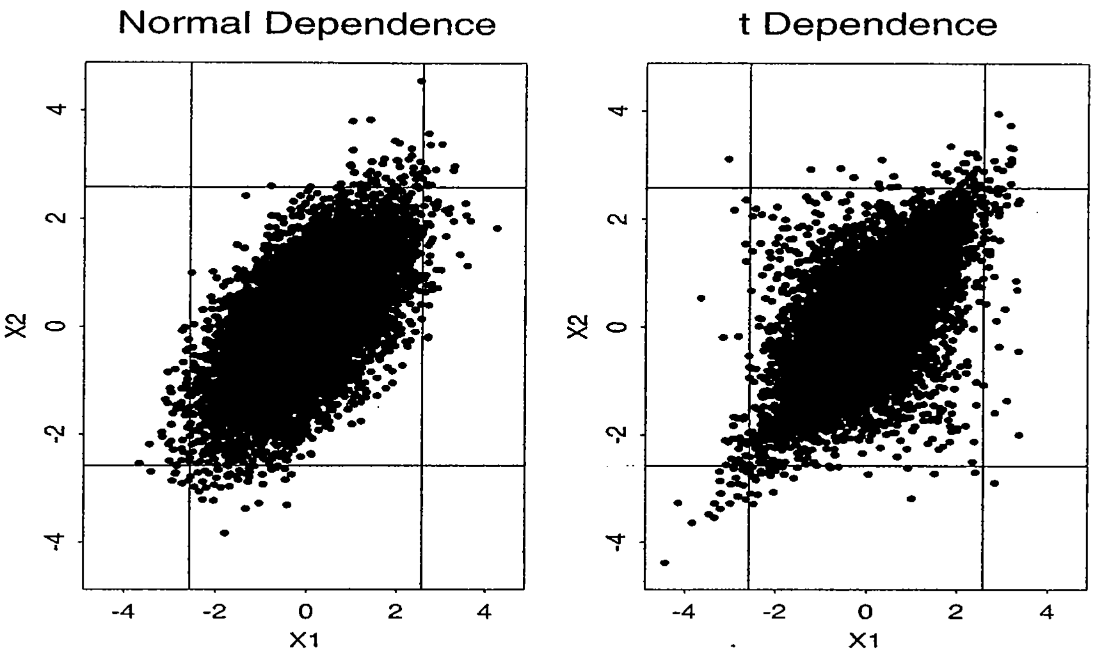
"Marginal distributions and correlation determine the joint distribution".

- True for the class of *bivariate* normal distributions or, more generally, for elliptical distributions.

- Wrong in general, as the next example shows.



Margins are standard normal; correlation is 70%.



Normal margins; correlation 70%; quantiles lines 0.5% and 99.5%.

Fallacy 1 continued

Sometimes Fallacy 1 is hidden in statements like:
"If two random variables X_1 and X_2 are uncorrelated, they may be considered as approximately independent".

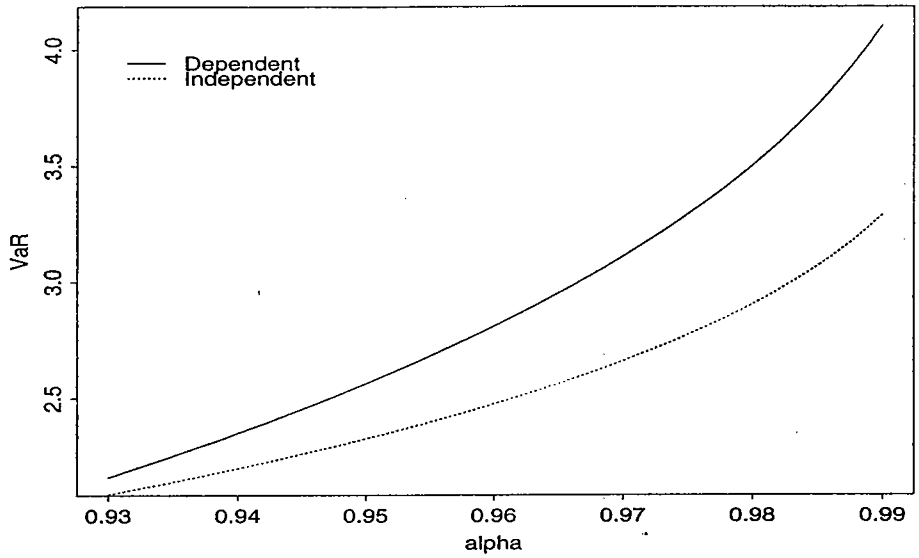
This view can be very dangerous in the management of risks. Consider two portfolios of risks. Set

$$X_1 = Z \quad \text{(Profit\&Loss Country A)},$$
$$X_2 = V \cdot Z \quad \text{(Profit\&Loss Country B)},$$

V, Z independent, $Z \sim N_d(\mathbf{0}, \mathrm{I})$,
$P(V = +1) = P(V = -1) = 1/2$.
V switches between perfect positive and negative dependence.

$X_2 \sim N_d(\mathbf{0}, \mathrm{I})$ and $\rho(X_1, X_2) = 0$.
But $(X_1, X_2)'$ is *not* bivariate normal.

VaR (Quantile) for two different dependence models



$\text{VaR}_\alpha(X_1 + X_2)$ for X_1, X_2 independent and X_1, X_2 dependent.

©2001 (Alexander McNeil and Rüdiger Frey)

Fallacy 2

"Given marginal distributions F_1 and F_2 for X_1 and X_2, all linear correlations between -1 and +1 can be attained through specification of the joint distribution".

- This is again true for elliptical distributions but not true in general. If F_1 and F_2 are not of the same type, then $\rho(X_1, X_2) < 1$.

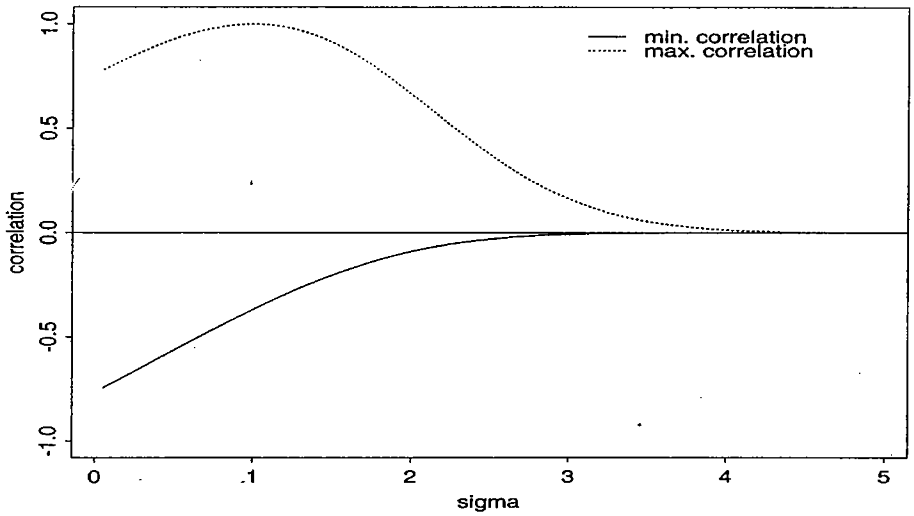
- **Theorem** (Höffding 1940)

  1. The set of possible correlations is a closed interval $[\rho_{\text{min}}, \rho_{\text{max}}]$.
  2. ρ_{max} is attained iff X_1, X_2 comonotonic. ρ_{min} is attained iff X_1, X_2 countermonotonic.

©2001 (Alexander McNeil and Rüdiger Frey)

Example of Extremal Correlations

Take $X_1 \sim \text{Lognormal}(0,1)$, and $X_2 \sim \text{Lognormal}(0,\sigma^2)$.
Let σ vary and plot $\rho_{\min}$ and $\rho_{\max}$ against σ.



3. Alternative Dependence Concepts

A Very Different Kind of Correlation: Rank Correlation

Spearman's rank correlation (Spearman's rho)

$$\rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)) = \rho(\text{copula}).$$

Kendall's rank correlation (Kendall's tau)
Take an independent copy of (X_1, X_2) denoted $(\widetilde{X}_1, \widetilde{X}_2)$.

$$\rho_\tau(X_1, X_2) = P\left((X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) > 0\right) - P\left((X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) < 0\right).$$

Suppose X_1 and X_2 have copula C. Then

$$\begin{aligned} \rho_\tau(X_1, X_2) &= 4\int_0^1\int_0^1 C(u_1,u_2)dC(u_1,u_2) - 1 \\ \rho_S(X_1, X_2) &= 12\int_0^1\int_0^1 \{C(u_1,u_2) - u_1u_2\}du_1du_2. \end{aligned}$$

Properties of Rank Correlation (not shared by linear correlation)

True for Spearman's rho (ρ_S) or Kendall's tau (ρ_τ).

- ρ_S depends only on copula of $(X_1, X_2)'$.
- ρ_S is invariant under strictly increasing transformations of the random variables.
- $\rho_S(X,Y) = 1 \iff X, Y$ comonotonic.
- $\rho_S(X,Y) = -1 \iff X, Y$ countermonotonic.

©2001 (Alexander McNeil and Rüdiger Frey)

Tail Dependence or Extremal Dependence

Objective: measure dependence in joint tail of bivariate distribution. When limit exists, coefficient of *upper* tail dependence is

$$\lambda_u(X_1, X_2) = \lim_{q \to 1} P(X_2 > \mathrm{VaR}_q(X_2) \mid X_1 > \mathrm{VaR}_q(X_1)),$$

Analogously the coefficient of *lower* tail dependence is

$$\lambda_\ell(X_1, X_2) = \lim_{q \to 0} P\left(X_2 \le \mathrm{VaR}_q(X_2) \mid X_1 \le \mathrm{VaR}_q(X_1)\right).$$

These are functions of the copula given by

$$\lambda_u = \lim_{q \to 1} \frac{\overline{C}(q,q)}{1-q} = \lim_{q \to 1} \frac{1 - 2q + C(q,q)}{1-q},$$
$$\lambda_\ell = \lim_{q \to 0} \frac{C(q,q)}{q}.$$

Thus they are invariant under strictly increasing transformations.

©2001 (Alexander McNeil and Rüdiger Frey)

Tail Dependence

Clearly $\lambda_u \in [0,1]$ and $\lambda_\ell \in [0,1]$.

For elliptical copulas $\lambda_u = \lambda_\ell =: \lambda$.
True of all copulas with $(U_1, U_2) \stackrel{d}{=} (1-U_1, 1-U_2)$ (known as radial symmetry) e.g. Frank copula.

Terminology:

$\lambda_u \in (0,1]$: upper tail dependence,
$\lambda_u = 0$: asymptotic independence in upper tail,
$\lambda_\ell \in (0,1]$: lower tail dependence,
$\lambda_\ell = 0$: asymptotic independence in lower tail.

©2001 (Alexander McNeil and Rüdiger Frey)

Examples of tail dependence

The Gaussian copula is asymptotically independent for $|\rho| < 1$.
The Frank copula is also asymptotically independent.

The t copula is tail dependent when $\rho > -1$.

$$\lambda = 2\bar{t}_{\nu+1}\left(\sqrt{\nu+1}\sqrt{1-\rho}/\sqrt{1+\rho}\right).$$

The Gumbel copula is upper tail dependent for $\beta > 1$.

$$\lambda_u = 2 - 2^{1/\beta}.$$

The Clayton copula is lower tail dependent for $\beta > 0$.

$$\lambda_\ell = 2^{-1/\beta}.$$

Recall dependence model in Fallacy 1b: $\lambda_u = \lambda_\ell = 0.5$.

©2001 (Alexander McNeil and Rüdiger Frey)

4. Bounds on VaR of Portfolios

Comonotonicity Revisited

Recall that the risks $X_1, \ldots, X_d$ are said to be comonotonic (perfectly positive dependent) if they have the copula

$$C(u_1, \ldots, u_d) = \min\{u_1, \ldots, u_d\}.$$

Essentially there is only one risk and all other risks are deterministic, increasing functions of this underlying risk.
In continuous case $X_i = T_i(X_1)$a.s., $i = 2, \ldots, d$, with T_i strictly increasing.

It may be shown that quantiles are additive for comonotonic risks:

$$\text{VaR}_\alpha\left(\sum_{i=1}^{d} X_i\right) = \sum_{i=1}^{d} \text{VaR}_\alpha(X_i).$$

However, it would be wrong to think this gives an upper bound for the risk of the sum in all cases. (If it did, VaR would be subadditive, and this is not true.)

Another Correlation Fallacy (or is this a VaR fallacy?)

"For fixed marginal distributions the worst case $\text{VaR}_\alpha(X_1 + X_2)$ occurs when $\rho(X_1, X_2)$ is maximal."

This seems intuitively plausible, but is only true for elliptical distributions. Let $Z = \lambda_1 X_1 + \lambda_2 X_2$ with (X_1, X_2) elliptical

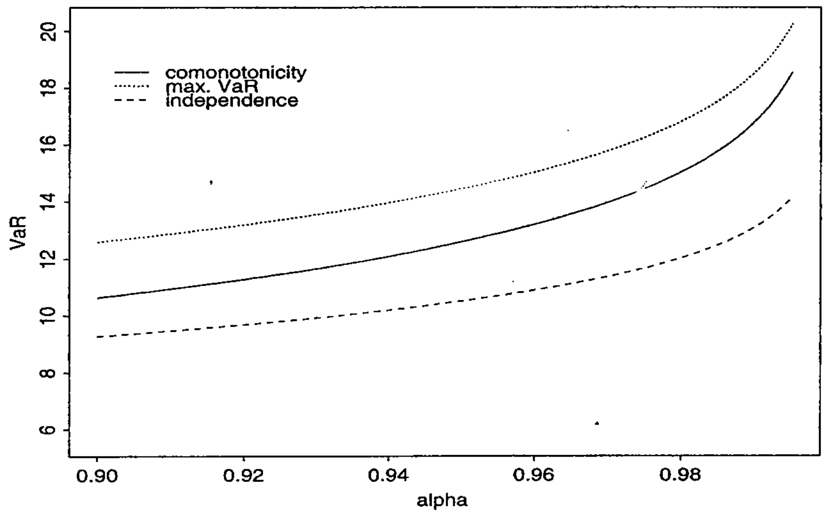
$$\text{VaR}_\alpha(Z) = E(Z) + \text{const}(\alpha) \cdot \text{sd}(Z),$$

and $\text{sd}(Z)$ is maximal when $\rho(X_1, X_2)$ is maximal.

In other cases one can calculate best-possible (pointwise) bounds for $\text{VaR}_\alpha(X_1 + X_2)$ (Frank et al. 1987).

Example

Take $X_1, X_2 \sim$ Gamma(3,1), and leave copula unspecified. Compute worst $\text{VaR}_\alpha(X_1 + X_2)$ against α.



References

On dependence in general:

- Joe (1997). Multivariate Models and Dependence Concepts. Chapman & Hall, London.

On copulas:

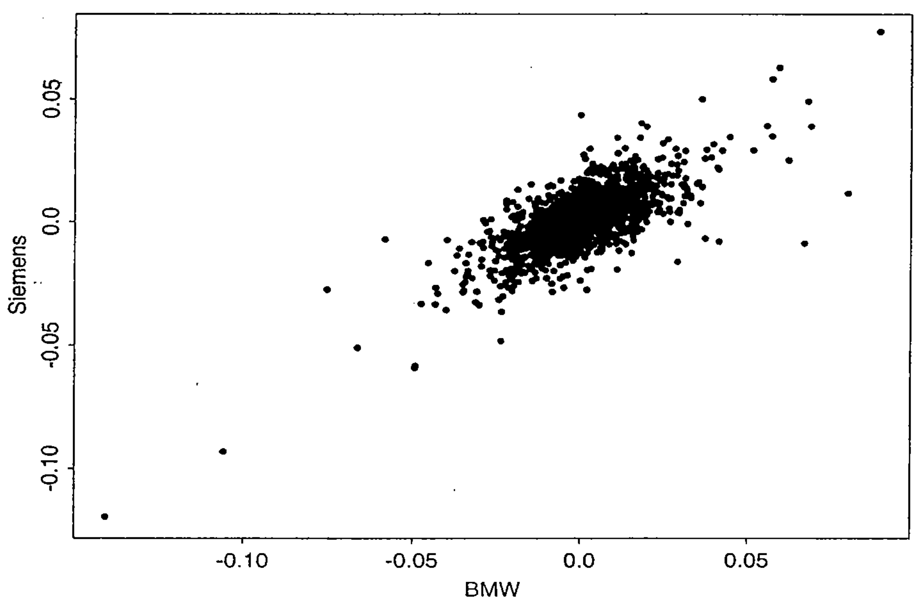
- Nelsen (1999). An Introduction to Copulas. Springer, New York
- Lindskog (2000). Modelling Dependence with Copulas. Preprint at http://www.math.ethz.ch/risklab

On correlation fallacies:

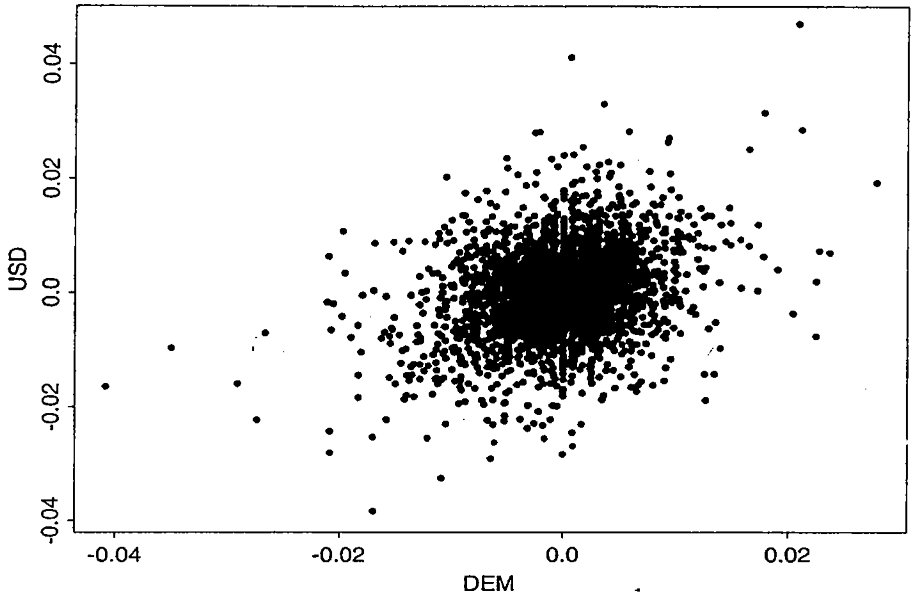
- Embrechts, McNeil & Straumann (2001). To appear in "Risk Management: Value at Risk and Beyond". CUP, Cambridge. Preprint at http://www.math.ethz.ch/~mcneil.

III. Calibration and Simulation of Multivariate Models

- Estimating correlation using standard methods
- Testing for non-multivariate-normal data
- Improved correlation estimation for heavy-tailed data
- Fitting copulas to data
- Generating dependent risk factors using Monte Carlo simulation
- Simulating meta-Gaussian distributions - the RISK method
- Simulating copulas



Scatterplot of Siemens and BMW returns on same days.



Scatterplot of Pound/DM and Pound/Dollar exchange rate returns.

©2001 (Alexander McNeil and Rüdiger Frey)

1. Estimating Correlation

Standard Approach to Estimating Correlations/Covariances

Suppose we have n observations of our d-dimensional risks: $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

Denote mean vector, covariance matrix and correlation matrix by μ, Σ and R respectively. Further notation:

$\overline{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i \quad Q = \sum_{i=1}^{n}(\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})'$ (componentwise).

The standard estimators used in practice are:

$$\begin{aligned}\widehat{\mu} &= \overline{\mathbf{X}} \\ \widehat{\Sigma}^{(1)} &= Q/n \quad \text{or} \quad \widehat{\Sigma}^{(2)} = Q/(n-1) \\ \widehat{R}_{ij} &= \widehat{\Sigma}_{ij}/\sqrt{\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}}\end{aligned}$$

©2001 (Alexander McNeil and Rüdiger Frey)

Properties of Estimators in Multivariate Normal Case

These standard estimators are designed *precisely* for the case when data are iid *multivariate normal*: $\mathbf{X}_i \sim N_d(\boldsymbol{\mu}, \Sigma)$.

In this case $\widehat{\boldsymbol{\mu}}$ and $\widehat{\Sigma}^{(1)}$ are the *maximum likelihood* estimates. $\widehat{\Sigma}^{(2)}$ is an alternative *unbiased* estimator.

Desirable estimator properties (such as *consistency* and *efficiency*) are well known.

However, we will *seldom* encounter multivariate normal data samples in finance.

Empirical Features of Risk Factor Time Series

(We think here particularly of market risk factors.)

Real multivariate risk factor time series

- may not be stationary (possible regime shifts)
- may show serial dependence (stochastic volatility effects)
- may not be multivariate normal (heavy tails - leptokurtosis)
- may not even be elliptical (lack of symmetry)

We want to assume at least stationarity - otherwise any kind of statistical inference is problematic. If regime really shifts we should look at data within regimes.

What Assumptions can be Relaxed?

Independence. $\widehat{\mu}$ and $\widehat{\Sigma}^{(2)}$ remain unbiased estimates in case where data are from stationary time series with multivariate normal stationary distribution. Properties depend on nature of serial dependence. If dependence is profound we should consider explicit time series modelling.

Normality. If data are really iid from some other elliptical distribution then $\widehat{\Sigma}$ and $\widehat{R}$ are no longer efficient. Other estimators should be preferred.

Ellipticality. If data are not elliptical then the wisdom of estimating correlations and covariances at all is called into question.

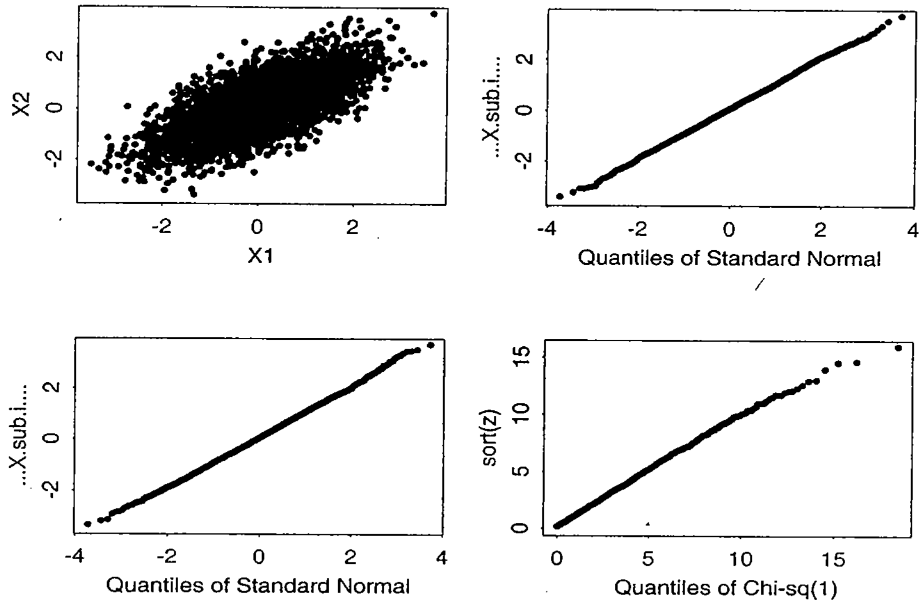
Testing for Multivariate Normality

If data are to be multivariate normal then margins must be univariate normal, but this is not sufficient - we require *joint* normality.

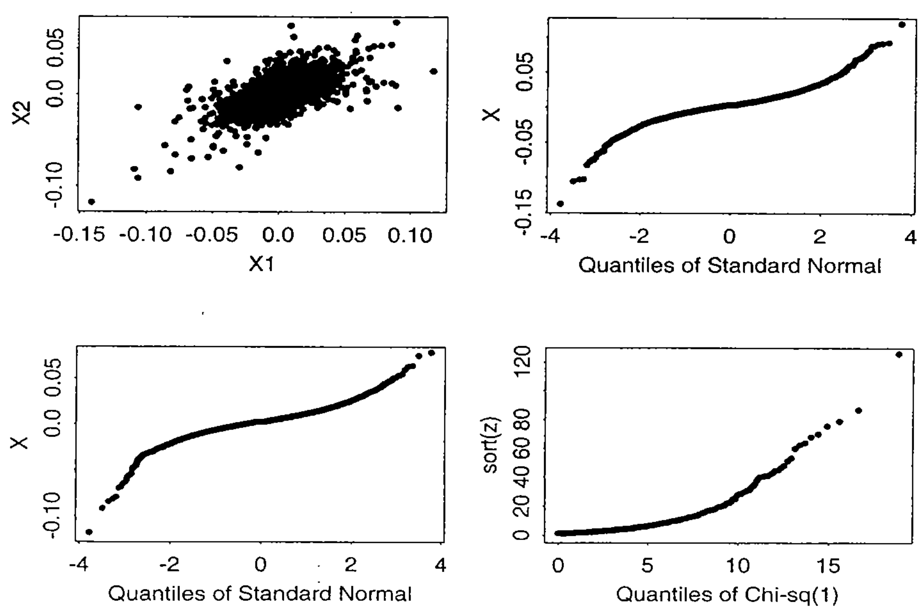
There are various numerical tests of normality, but we will concentrate here on appealing visual tests such as the QQplot.
(QQplots compare empirical quantiles with theoretical quantiles of reference distribution.)

To test joint normality calculate $\{(\mathbf{X}_i - \widehat{\mu})'\widehat{\Sigma}^{-1}(\mathbf{X}_i - \widehat{\mu}),\ i = 1, \ldots, n\}$; these should form a sample from a χ^2_d-distribution, which can be assessed with a QQplot.

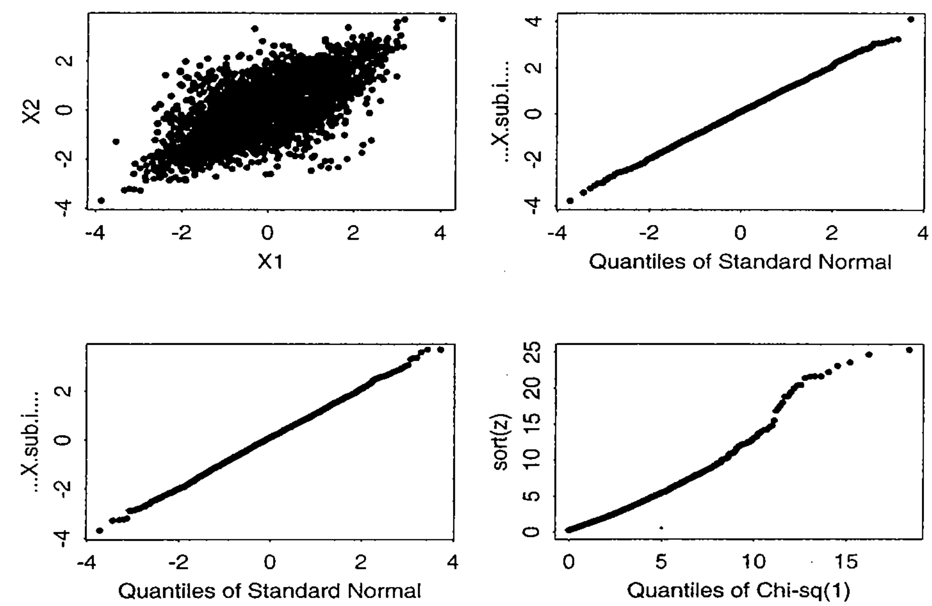
Testing Multivariate Normality: Normal Data



Testing Multivariate Normality: BMW-Siemens Data



Testing Multivariate Normality: Normal Margins - t4 copula



Testing for Ellipticality

If $\mathbf{X}$ has an elliptical distribution with mean vector μ, finite covariance matrix Σ and correlation matrix R, then

$$\text{corr}\left(\mathbf{X} \mid (\mathbf{X}-\mu)'\Sigma^{-1}(\mathbf{X}-\mu) > l\right) = R, \quad \forall l > 0.$$

This suggests a graphical method:

1. Estimate μ and Σ.
2. For fixed d, select points lying outside the ellipsoid defined by
$$(\mathbf{X}-\widehat{\mu})'\widehat{\Sigma}^{-1}(\mathbf{X}-\widehat{\mu}) = l.$$
3. Estimate R using points outside ellipse.
4. Repeat for various d and assess stability of correlations.

Efficient Covariance/Correlation Estimation for Elliptical Data

Assume we have iid data from an elliptical distribution. In seeking alternative estimates there are two considerations.

Efficiency
We seek estimators whose *expected distance* from the true covariance/correlation matrix is as small as possible. One estimator is said to be more efficient than an other if its expected distance from the true matrix is always smaller.

Robustness
A covariance/correlation estimator is robust if it is insensitive to *contamination* of the elliptical population by small numbers of *outliers* from another population.
(Obviously we seek estimators with a certain amount of robustness. But if a large contamination of the elliptical population is apparent we should really be trying to model this rather than mitigate its effect on correlation estimation.)

A. M-Estimators of Covariance/Correlation (Maronna 1976)

Intuition: downweight observations which are large compared to average.

1. Take $\widetilde{\mu}$ and $\widetilde{\Sigma}$ to be $\widehat{\mu}$ and $\widehat{\Sigma}$.

2. Set $M_i^2 = (\mathbf{X}_i - \widetilde{\mu})'\widetilde{\Sigma}^{-1}(\mathbf{X}_i - \widetilde{\mu}), \quad i = 1, \ldots, n.$

3. Set
$$\widetilde{\mu} = \frac{\sum_{i=1}^n w_1(M_i)\mathbf{X}_i}{\sum_{i=1}^n w_1(M_i)} \quad \widetilde{\Sigma} = \frac{1}{n}\sum_{i=1}^n w_2(M_i^2)(\mathbf{X}_i - \widetilde{\mu})(\mathbf{X}_i - \widetilde{\mu})',$$
where w_1 and w_2 are weight functions.

4. Calculate $\widetilde{R}^{\mathrm{M}}$. If $\widetilde{R}^{\mathrm{M}}$ is *sufficiently stable* stop, otherwise go to 2.

The following popular choice of w_1 and w_2 essentially gives maximum-likelihood estimators for the multivariate t_ν distribution:
$w_1(x) = (d + \nu)/(x^2 + \nu) = w_2(x^2)$.

B. An Estimator Based on Rank Correlation

Estimating Rank Correlations

An estimator of ρ_τ is given by

$$\widehat{\rho_\tau}(X_1, X_2) = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \mathrm{sgn}[(X_{1,i} - X_{1,j})(X_{2,i} - X_{2,j})].$$

Spearman's rank correlation ρ_S is estimated by

$$\widehat{\rho_S}(X_1, X_2) = \frac{12}{n(n^2-1)} \sum_{i=1}^{n} \left(\mathrm{rank}(X_{1,i}) - \frac{n+1}{2}\right)\left(\mathrm{rank}(X_{2,i}) - \frac{n+1}{2}\right).$$

Estimating Elliptical Correlations: Kendall's tau transform

Another efficient and robust estimator of correlation may be based on the following result. With this estimator covariance need not be estimated.

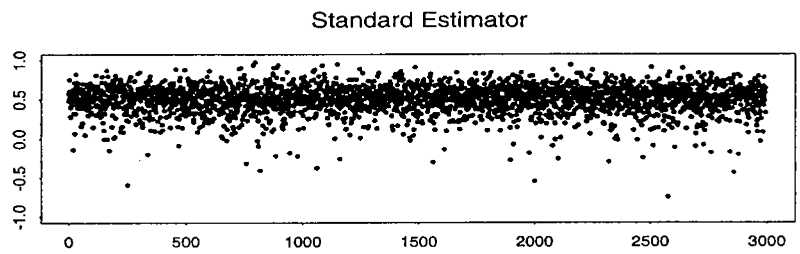
Proposition. For (essentially) all elliptical distributions we have $\rho_\tau(X_i, X_j) = \frac{2}{\pi}\arcsin(\rho(X_i, X_j))$

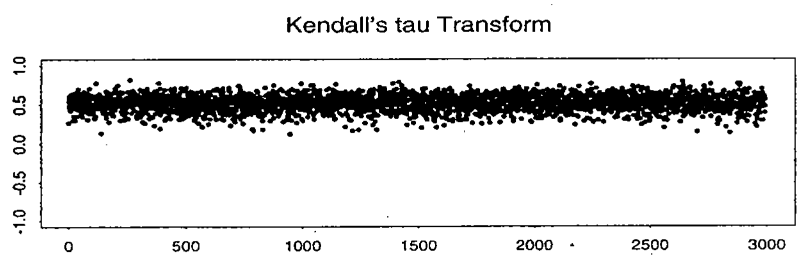
1. Estimate pairwise Kendall's rank correlations $\widehat{\rho_\tau}_{ij}$.

2. Calculate pairwise linear correlation estimates according to $\widehat{\rho}^{\tau}_{ij} = \sin\left(\frac{\pi}{2}\widehat{\rho_\tau}_{ij}\right)$.

3. Check that resulting estimate of correlation matrix $\widehat{R}^\tau$ is positive definite. If not, adjust elements slightly with *eigenvalue method* (Lindskog 2000).
Our experience in a simulation with $d = 4$ suggests that $\widehat{R}^\tau$ will be positive definite over 90% of the time.

Efficient Correlation Estimation with Kendall's Tau Transform





Simulated t_3 data; 3000 times 90 pairs; true value 0.5

©2001 (Alexander McNeil and Rüdiger Frey)

2. Fitting Copulas to Data: An Example

We fit *bivariate* copulas to real data. The method extends to higher dimensions, but seems mot suitable for low dimensions - say ≤ 4,

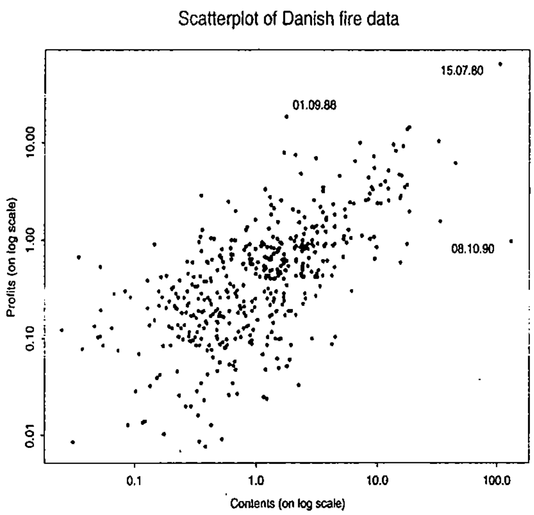
The data comprise $n = 466$ Danish fire insurance losses decomposed into two components: $X_1 =$ `Loss of Contents`, $X_2 =$ `Loss of Profits`.

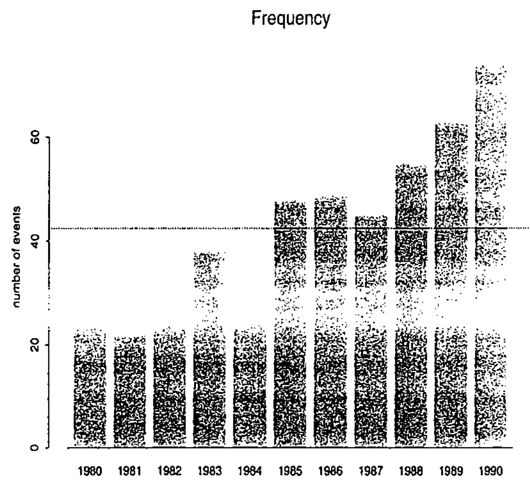
Obviously data in this format could be collected on various kinds of financial loss: losses in two (or more) lines of business; credit and market risk losses.

We adopt a *two-stage procedure*:
1. Fit distributions to the marginals;
2. Estimate the "dependence", i.e. the copula.

©2001 (Alexander McNeil and Rüdiger Frey)

The Data





©2001 (Alexander McNeil and Rüdiger Frey)

Estimating Copulas via MLE

We consider four parametric copulas, each with a single parameter: Gumbel, Clayton, Frank and Gaussian.

Denote data $\{(X_{1,i}, X_{2,i}), i = 1, \ldots, n\}$.
The copula data $\{(F_1(X_{1,i}), F_2(X_{2,i})), i = 1, \ldots, n\}$ are not themselves directly observable since F_1 and F_2 are unknown.

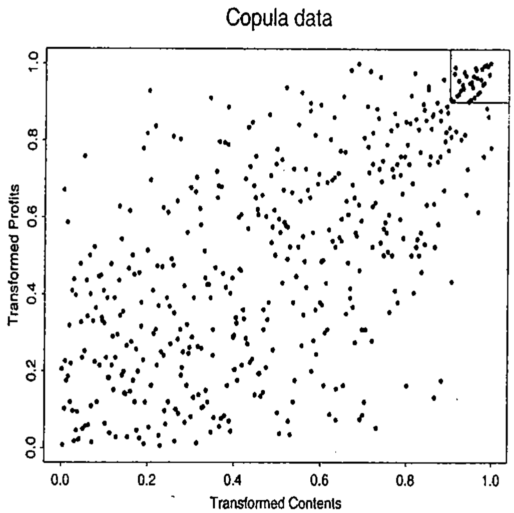
Stage 1. Estimate margins using version of empirical distribution function:

$$\widehat{F}_1(x) = \frac{1}{n+1}\sum_{i=1}^{n} 1_{\{X_{1,i}\le x\}} \quad \widehat{F}_2(x) = \frac{1}{n+1}\sum_{i=1}^{n} 1_{\{X_{2,i}\le x\}}$$

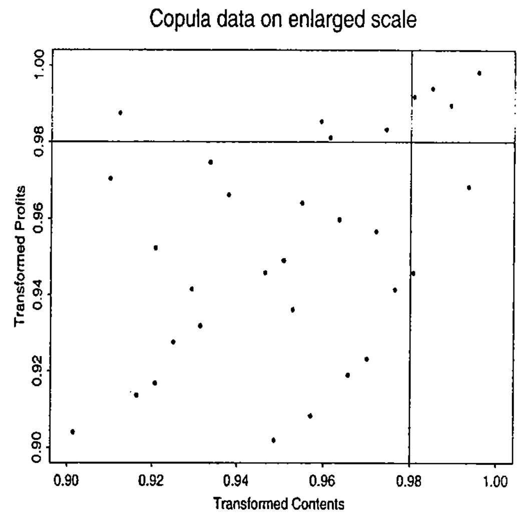
and apply the MLE method to transformed data
$\{(\widehat{F}_1(X_{1,i}), \widehat{F}_2(X_{2,i})), i = 1, \ldots, n\}$.

©2001 (Alexander McNeil and Rüdiger Frey)

Transformed Data



Tail Dependence



$\widehat{\lambda}(X_1, X_2) = 4/10 = 0.4$ (crude).

Stage 2: Estimated Parameters

Copula	β	std.error	neg. log-likelihood
Gumbel	1.85	0.07	−142.27
Clayton	0.85	0.09	−66.71
Frank	5.09	0.34	−123.50
Gaussian	0.65	0.02	−126.67

Goodness-of-fit

Akaike's information criterion (**AIC**) suggests choosing model which minimizes

$$\text{AIC} = 2 \cdot (\text{neg. log-likelihood}) + 2p,$$

where $p =$ number of model parameters. This would be the Gumbel model.

Comment: Graphical methods for assessing fit also available.

Copulas at Work: Using the Fitted Model

We now have a model in the form of the distribution function

$$\widehat{F}(x_1, x_2) = \widehat{C}(\widehat{F}_1(x_1), \widehat{F}_2(x_2)),$$

where $\widehat{C}$ is the fitted copula,and $\widehat{F}_1$ and $\widehat{F}_2$ are fitted marginal distributions (either the empirical dfs or univariate parametric models).

We can now evaluate the distribution of loss functions $L(X_1, X_2)$ depending on X_1 and X_2 (often via simulation).

We consider two payout functions:

$$\begin{aligned} L_1(X_1, X_2) &= (X_1 + X_2 - 10)_+, \\ L_2(X_1, X_2) &= X_2 \cdot 1_{\{X_1 \geq 10\}} \end{aligned}$$

and estimate the *expected* loss $P_1 = E(L_1), P_2 = E(L_2)$.

Estimated Expected Losses for 2 Payout Functions

We estimate P_1 and P_2 using a Monte Carlo procedure, i.e. we simulate from the distribution of L_1 and L_2 (1 mio. runs).

Copula	$\widehat{P}_1$	$\widehat{\sigma}(\widehat{P}_1)$
Gumbel	1.28	0.01
Clayton	1.11	0.01
Frank	1.18	0.01
Normal	1.24	0.01
Independence	1.05	0.01

Copula	$\widehat{P}_2$	$\widehat{\sigma}(\widehat{P}_2)$
Gumbel	0.094	< 0.001
Clayton	0.072	< 0.001
Frank	0.098	< 0.001
Normal	0.094	< 0.001
Independence	0.049	< 0.001

3. Multivariate Simulation

Simulating Normal-Mixture Distributions

It is straightforward to simulate normal mixture models. We only have to simulate a Gaussian random vector and an independent radial random variable. Simulation of Gaussian vector in all standard texts.

Example: t distribution
To simulate a vector $\mathbf{X}$ with distribution $t_d(\nu, \boldsymbol{\mu}, \Sigma)$ we would simulate $\mathbf{Z} \sim N_d(\boldsymbol{\mu}, \Sigma)$ and $V \sim \chi^2_\nu$;
we would then set $W = \sqrt{\nu/V}$ and $\mathbf{X} = \boldsymbol{\mu} + W\mathbf{Z}$.

To simulate *generalized hyperbolic* distributions we are required to simulate a radial variate with the NIG distribution. See Atkinson (1982) for an algorithm; see also work of Eberlein and Prause.

Simulating Copulas of Normal Mixtures
(particularly elliptical copulas)

It is thus also straightforward to simulate from the copulas of normal-mixture models.

To simulate from the Gaussian copula C_R^{Ga} we would:
1. Simulate $\mathbf{X} \sim N_d(\mathbf{0}, R)$
2. Probability transform margins: $\mathbf{U} = (\Phi(X_1), \ldots, \Phi(X_d))'$.

For example, to simulate from the t copula $C_{\nu,R}^{\mathrm{t}}$ we would:
1. Simulate $\mathbf{X} \sim t_d(\nu, \mathbf{0}, R)$
2. Probability transform margins: $\mathbf{U} = (t_\nu(X_1), \ldots, t_\nu(X_d))'$,
where t_ν is df of univariate t distribution.

Remark. To simulate $\mathbf{X} \sim N_d(\mathbf{0}, R)$ use Cholesky's decomposition to get $R = AA'$ and simulate $\mathbf{Z} \sim N_d(\mathbf{0}, \mathrm{I})$. Then $\mathbf{X} = A\mathbf{Z}$ has required distribution.

Meta-Gaussian and Meta-t Distributions

If $(U_1, \ldots, U_d) \sim C_R^{\mathrm{Ga}}$ and G_i are univariate dfs other than univariate normal then

$$\left(G_1^{-1}(U_1), \ldots, G_d^{-1}(U_d)\right)$$

has a *meta-Gaussian* distribution. Thus it is easy to simulate vectors with the Gaussian copula and arbitrary margins.

In a similar way we can construct and simulate from *meta t_ν distributions*. These are distributions with copula $C_{\nu,R}^{\mathrm{t}}$ and margins other than univariate t_ν.

A Simulation Problem from Practice

Let $F_1, \ldots, F_d$ be univariate distributions and let R be a positive definite correlation matrix with i,jth element ρ_{ij}.

Simulate random vector of risks $(X_1, \ldots, X_d)$ from multivariate model such that

1. $$X_i \sim F_i, \quad i = 1, \ldots, d,$$

2a. $$\rho(X_i, X_j) = \rho_{ij}, \quad i,j = 1, \ldots, d.$$

But is this a well-posed problem?
What if we replace 2 by

2b. $$\rho_S(X_i, X_j) = \rho_{ij}, \quad i,j = 1, \ldots, d\ ?$$

The computer program @RISK purports to solve this second problem.

Simulation of random vectors with given marginals and correlation matrix

(1,2a) An *ill-posed* problem. We must be aware of consistency and uniqueness problems. There may be no solutions or there may be infinitely many solutions. (Fallacies 1 and 2.) There are however ad–hoc methods to construct joint distributions with given marginals and correlation matrix when a solution exists.

(1,2b) If a matrix of rank correlations is given the situation is *better,* since Fallacy 2 is avoided. However Fallacy 1 remains and the problem is still ill-posed. An approach by Iman and Conover based on simulating from a *Gaussian copula* is implemented in the @RISK computer programme.

©2001 (Alexander McNeil and Rüdiger Frey)

The @RISK Method

Variant 1.
Simulate $\widetilde{\mathbf{X}} \sim N_d(\mathbf{0}, R)$ and set

$$(X_1, \ldots, X_d) = \left(F_1^{-1}(\Phi(\widetilde{X_1})), \ldots, F_d^{-1}(\Phi(\widetilde{X_d}))\right).$$

This is an approximate solution since

$$\rho_S(X_i, X_j) = \rho_S(\widetilde{X}_i, \widetilde{X}_j) = \frac{6}{\pi} \arcsin \frac{\rho(\widetilde{X}_i, \widetilde{X}_j)}{2} \approx \rho_{ij}.$$

Variant 2.
Simulate $\widetilde{\mathbf{X}} \sim N_d(\mathbf{0}, \tilde{R})$ where elements of $\tilde{R}$ satisfy $\tilde{\rho}_{ij} = 2 \sin \frac{\pi \rho_{ij}}{6}$. Set

$$(X_1, \ldots, X_d) = \left(F_1^{-1}(\Phi(\widetilde{X_1})), \ldots, F_d^{-1}(\Phi(\widetilde{X_d}))\right).$$

If $\tilde{R}$ is positive definite then this is an exact solution.

Drawback: we only use the Gaussian copula.

©2001 (Alexander McNeil and Rüdiger Frey)

Simulating Other Copulas: A few Comments

For parametric copulas a variety of simulation techniques exist, which are suitable for various families such as the Archimedean. See Lindskog (2000) for details.

Useful algorithm for simulating (U_1, U_2) from bivariate copula $C(u_1, u_2)$. Let $C_2(u_2 \mid u_1) = \frac{d}{du_1} C(u_1, u_2)$. This is conditional distribution of U_2 given $U_1 = u_1$.
1. Simulate $U_1 \sim U(0,1)$
2. Simulate U_2 from distribution with df $C_2(u_2 \mid u_1)$.

This idea may be extended to higher dimensions - see Embrechts, McNeil, Straumann (2001).

©2001 (Alexander McNeil and Rüdiger Frey)

References

On efficient and robust correlation estimation in elliptical models:

- Lindskog (2000). Linear correlation estimation. Preprint at `http://www.math.ethz.ch/risklab`

On fitting copulas to data:

- Klugman & Parsa (1999). Fitting bivariate loss distributions with copulas. Insurance: Mathematics and Economics 24.
- Frees & Valdez (1997). Understanding relationships using copulas. N. American Actuarial Journal 2,1.

On multivariate simulation:

- Lindskog (2000). Modelling dependence with copulas. Preprint at `http://www.math.ethz.ch/risklab`

©2001 (Alexander McNeil and Rüdiger Frey)

IV. Multivariate Models and Copulas for Credit Risk

- Multivariate discrete models for credit risk
- Latent variables models and mixture models
- Examining standard solutions: CreditMetrics, KMV, CreditRisk+
- Mapping between latent variable and mixture models
- What is extreme credit risk?
- Copulas and extreme credit risk
- Improving and extending standard solutions
- Generating risky scenarios - a simulations study
- Alternative risk transfer - basket credit derivatives
- Calibrating credit models to available information

©2001 (Alexander McNeil and Rüdiger Frey)

Motivation

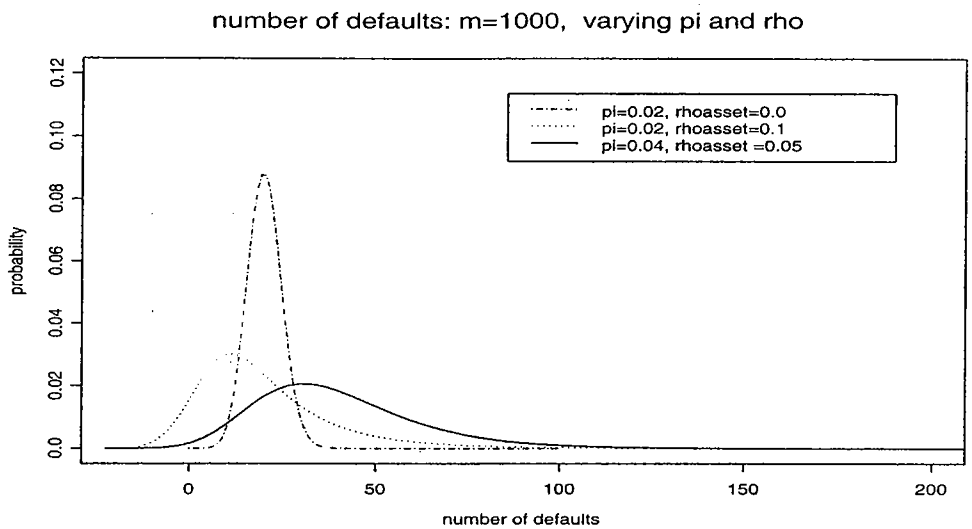
- Focus in credit risk research has mainly been on modelling of default of individual firm (e.g. firm-value models versus reduced-form models, rating class versus asset value of firm as state variable for default probability).
- Modelling of joint defaults in standard models (KMV, CreditMetrics) is relatively simplistic (based on multivariate normality).
- In large balanced loan portfolios main risk is occurrence of many joint defaults - this might be termed *extreme credit risk*.

Aim

Reconsider modelling of dependent defaults from viewpoint of recent research on "Correlation and Dependence in Risk Management."

©2001 (Alexander McNeil and Rüdiger Frey)

Dependent defaults and credit losses



Distribution of number of defaults in portfolio of 1000 firms.
Dependence between defaults has a large influence on distribution.

©2001 (Alexander McNeil and Rüdiger Frey)

Modelling of Default – Overview

Consider portfolio of m firms and a time-horizon T (typically 1 year). For $i \in \{1, \ldots, m\}$ define Y_i to be default indicator of company i, i.e. $Y_i = 1$ if company defaults by time T, $Y_i = 0$ otherwise. (Reduction to two states (default/no default) for simplicity.)

Model Types

- *Latent variable models*
  Default occurs, if a latent variable X_i (often interpreted as asset value at horizon T) lies below some threshold D_i (liabilities).
  Examples: Merton model (1974), CreditMetrics, KMV.

- *Mixture Models*
  Bernoulli default probabilities are made stochastic.
  $Y_i \mid Q_i \sim Be(Q_i)$ where Q_i is a random variable taking values in $[0, 1]$ and $Q_1, \ldots, Q_m$ are dependent.
  Example: CreditRisk$^+$.

©2001 (Alexander McNeil and Rüdiger Frey)

1. Latent Variable Models

Given random vector $\mathbf{X} = (X_1, \ldots, X_m)'$ with continuous marginal distributions F_i and thresholds $D_1, \ldots, D_m$, define $Y_i := 1_{\{X_i \le D_i\}}$. *Default probability* of counterparty i given by

$$p_i := P(Y_i = 1) = P(X_i \le D_i) = F_i(D_i).$$

Notation: $(X_i, D_i)_{1 \le i \le m}$ denotes a latent variable model.

Examples

- *Classical Merton-model.*
  X_i is interpreted as asset value of company i at T. D_i is value of liabilities. Assume $\mathbf{X} \sim N(\mu, \Sigma)$.

Industry Examples of Latent Variable Models

- *KMV-model.*
  As Merton but D_i is now chosen so that default probability p_i equals average default probability of companies with same "distance to default" as company i.

- *CreditMetrics.*
  We assume $\mathbf{X} \sim N(0, \Sigma)$. Threshold D_i is chosen so that p_i equals average default probability of companies with same rating class as company i.

- *Model of Li* (CreditMetrics Monitor 1999)
  X_i interpreted as survival time of company i.
  Assume X_i exponentially distributed with parameter λ chosen so that $P(X_i \le T) = p_i$, with p_i determined as in CreditMetrics.
  Multivariate distribution of $\mathbf{X}$ specified using Gaussian copula.

Model Calibration (KMV and CreditMetrics)

In both KMV and CreditMetrics, μ_i, Σ_{ii} and D_i are chosen so that p_i equals average historical default frequency for companies with a *similar credit quality.* (They use different methods to group companies by credit quality.)

To determine further structure of Σ (i.e. correlations) both models assume a classical *linear factor model* for $p < m$.

$$X_i = \mu_i + \sum_{j=1}^{p} a_{i,j}\Theta_j + \sigma_i \varepsilon_i$$

for a p-dimensional random vector $\mathbf{\Theta} \sim N_p(\mathbf{0}, \Omega)$, *independent* standard normally distributed rv's $\varepsilon_1, \ldots, \varepsilon_m$, which are also *independent* of $\mathbf{\Theta}$.

- $\mathbf{\Theta}$ global, country and industry effects impacting all companies.
- $a_{i,j}$ loadings or weights for company i and factor j.
- ϵ idiosyncratic effects.

Equivalent Latent Variable Models and Copulas

Definition: Two latent variable models $(X_i, D_i)_{1\leq i\leq m}$ and $(\widetilde{X}_i, \widetilde{D}_i)_{1\leq i\leq m}$ generating multivariate Bernoulli vectors $\mathbf{Y}$ and $\widetilde{\mathbf{Y}}$ are said to be *equivalent* if $\mathbf{Y} \stackrel{\mathrm{d}}{=} \widetilde{\mathbf{Y}}$.

Proposition
$(X_i, D_i)_{1\leq i\leq m}$ and $(\widetilde{X}_i, \widetilde{D}_i)_{1\leq i\leq m}$ are equivalent if:

1. $P(X_i \leq D_i) = P(\widetilde{X}_i \leq \widetilde{D}_i)$, $\forall i$.
2. $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ have the same *copula*.

CreditMetrics and KMV are equivalent, as are all latent variable models that use the Gaussian dependence structure for latent variables, such as the model of Li, regardless of how marginals are modelled. .

Special Case: Homogeneous Groups - Exchangeability

It is common to group obligors together to form homogeneous groups. This corresponds to the mathematical concept of *exchangeability*. A random vector $\mathbf{X}$ is exchangeable if

$$(X_1, \dots, X_m) \stackrel{\mathrm{d}}{=} (X_{p(1)}, \dots, X_{p(m)}),$$

for any permutation $(p(1), \dots, p(m))$ of $(1, \dots, m)$.

We talk of an *exchangeable default model* if the default indicator vector $\mathbf{Y}$ is exchangeable.
If a latent variable vector $\mathbf{X}$ is exchangeable (or has an exchangeable copula) and all individual default probabilities $P(X_i \le D_i)$ are equal, then $\mathbf{Y}$ is exchangeable.

Exchangeability allows a simplified notation for default probabilities:

$$\begin{aligned} \pi_k &:= P(Y_{i_1} = 1, \dots, Y_{i_k} = 1), \quad \{i_1, \dots, i_k\} \subset \{1, \dots, m\},\ 1 \le k \le m, \\ \pi &:= \pi_1 = P(Y_i = 1), \quad i \in \{1, \dots, m\}. \end{aligned}$$

©2001 (Alexander McNeil and Rüdiger Frey)

The Copula is Critical

To see this consider special case of exchangeable default model.

Consider *any* subgroup of k companies $\{i_1, \dots, i_k\} \subset \{1, \dots, m\}$.

$$\begin{aligned} \pi_k = P(Y_{i_1} = 1, \dots, Y_{i_k} = 1) &= P\left(X_{i_1} \le D_{i_1}, \dots, X_{i_k} \le D_{i_k}\right) \\ &= C_{1,\dots,k}(\pi, \dots, \pi), \end{aligned}$$

where $C_{1,\dots,k}$ is the k-dimensional margin of C.

The copula C crucially determines higher order joint default probabilities and thus *extreme risk* that many companies default. For π small, copulas with lower tail dependence will lead to higher π_k's and more joint defaults.

©2001 (Alexander McNeil and Rüdiger Frey)

Comparison of Exchangeable Gaussian and t Copulas

If $\mathbf{X}$ is given an asset value interpretation large (downward) movements of the X_i might be expected to occur together; therefore tail dependence may be realistic.

We concentrate on two cases (extensions such as generalized hyperbolic distributions can be considered analogously).

1. $\mathbf{X} \sim N_m(\mathbf{0}, R)$
2. $\mathbf{X} \sim t_{m,\nu}(\mathbf{0}, R)$.

R is an equicorrelation matrix with off-diagonal element $\rho > 0$, so that $\mathbf{X}$ is exchangeable with correlation matrix R in both cases.

We also fix thresholds so that $\mathbf{Y}$ is exchangeable in both cases and $P(Y_i = 1) = \pi$, $\forall i$, in both models. We choose a value for ν.
How do the two models compare?

©2001 (Alexander McNeil and Rüdiger Frey)

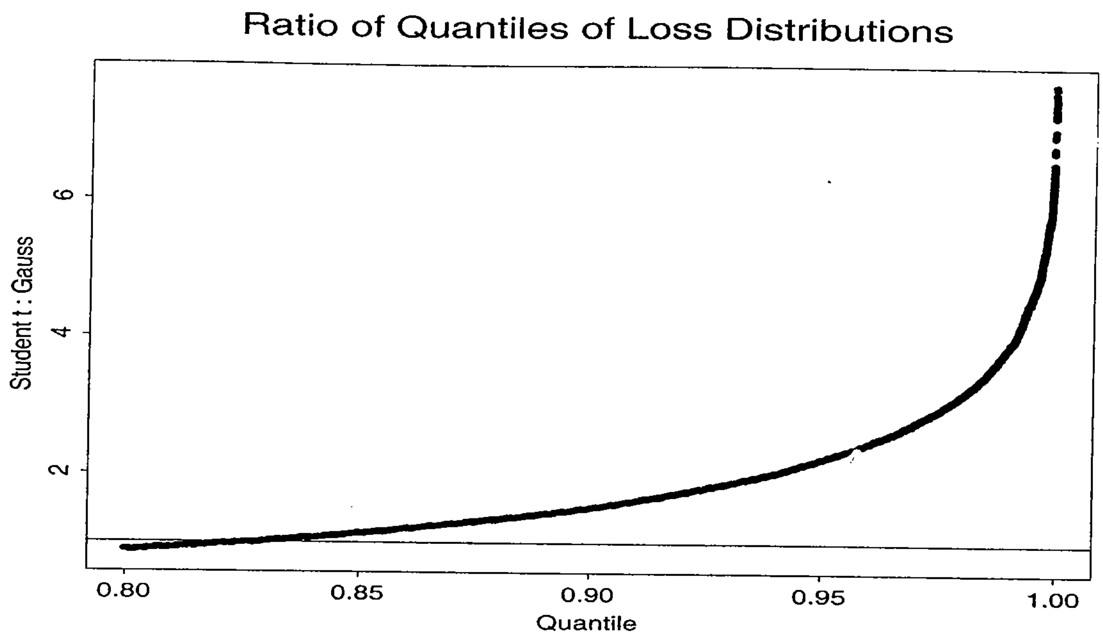
Simulation Study

We consider $m = 10000$ companies. All losses given default are one unit; total loss is number of defaulting companies.
Set $\pi = 0.005$ and $\rho = 0.038$, these being values corresponding to a homogeneous group of "medium" credit quality in the KMV/CreditMetrics Gaussian approach.
We set $\nu = 10$ in t-model and perform 100000 simulations to determine loss distribution.

The risk is compared by comparing high quantiles of the loss distributions (the so-called Value-at-Risk approach to measuring risk).

Results	Min	25%	Med	Mean	75%	90%	95%	Max
Gauss	1	28	43	49.8	64	90	109	331
t	0	1	9	49.9	42	132	235	3238

©2001 (Alexander McNeil and Rüdiger Frey)



Ratio of quantiles of loss distributions (t:Gaussian).
$m = 10000$, $\pi = 0.005$, $\rho = 0.038$ and $\nu = 10$.

2. Exchangeable Bernoulli Mixture Models

(We start with exchangeable case before generalising.)

The default indicator vector $(Y_1, \ldots, Y_m)$ follows an exchangeable Bernoulli mixture model if there exists a rv Q taking values in $(0,1)$ such that, given Q, $Y_1, \ldots, Y_m$ are iid Be(Q) rvs.

In such a model

$$\begin{aligned} \pi &= P(Y_i = 1) = E(Y_i) = E\left(E(Y_i \mid Q)\right) = E(Q) \\ \pi_k &= P(Y_{i_1} = 1, \ldots, Y_{i_k} = 1) = E(Q^k) = \int_0^1 q^k dG(q), \end{aligned}$$

where $G(q)$ is the *mixture distribution function* of Q. Unconditional default probabilities and higher order joint default probabilities are moments of the mixing distribution.

It follows that, for $i \neq j$, $\text{cov}(Y_i, Y_j) = \pi_2 - \pi^2 = \text{var}Q \geq 0$.
Default correlation is given by $\rho_Y := \text{corr}(Y_i, Y_j) = \frac{\pi_2 - \pi^2}{\pi - \pi^2}$.

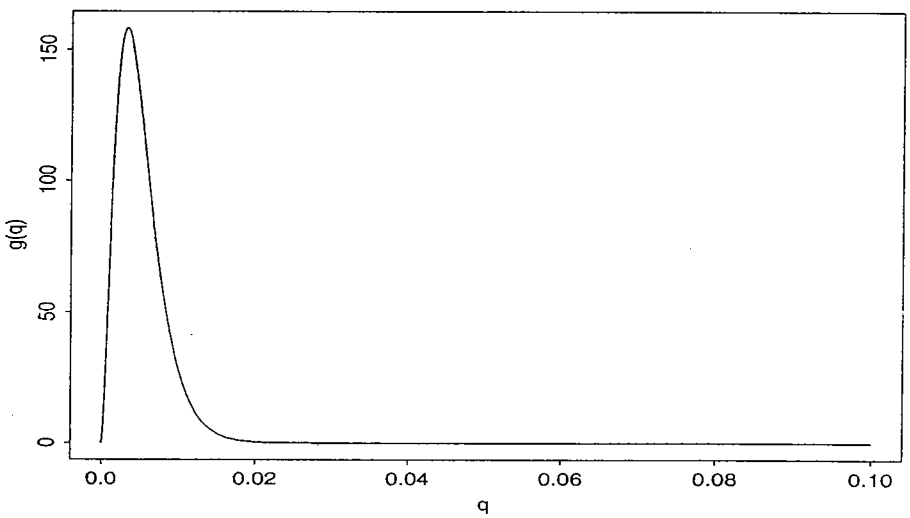
Examples of Mixing Distributions

- *Beta* $Q \sim \text{Beta}(a, b)$, $g(q) = \beta(a,b)^{-1} q^{a-1}(1-q)^{b-1}$, $a, b > 0$
- *Probit-Normal* $\Phi^{-1}(Q) \sim N(\mu, \sigma^2)$ (CreditMetrics/KMV)
- *Logit-Normal* $\log\left(\frac{Q}{1-Q}\right) \sim N(\mu, \sigma^2)$ (CreditPortfolioView)

Parameterizing Mixing Distributions
These examples all have two parameters. If we fix the default probability π and default correlation ρ_Y (or equivalently the first two moments of the mixing distribution π and π_2) then we fix these two parameters and fully specify the model.

Example: Exchangeable Beta-Bernoulli Mixture Model
$\pi = a/(a+b)$, $\pi_2 = \pi(a+1)/(a+b+1)$.

Beta Mixing Distribution



Beta Density $g(q)$ of mixing variable Q in exchangeable Bernoulli mixture model with $\pi = 0.005$ and $\rho_Y = 0.0018$.

Extreme Credit Risk in Exchangeable Mixture Models

Recall that in exchangeable latent variable models

$$\pi_k = C_{1,\ldots,k}(\underbrace{\pi,\ldots,\pi}_{k \text{ times}}),$$

and in exchangeable Bernoulli mixture models

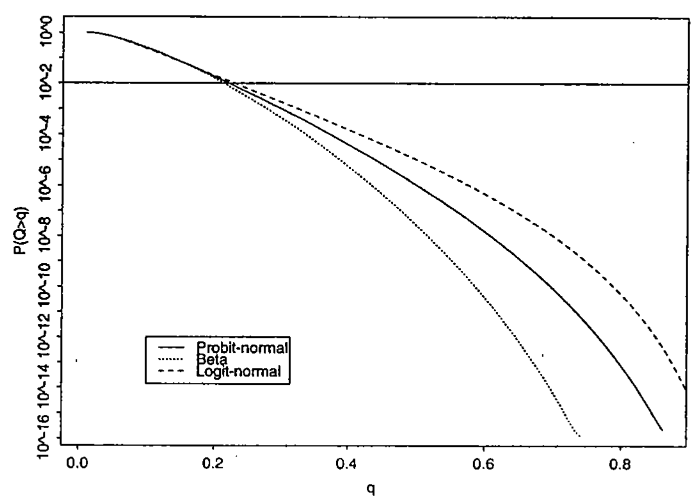
$$\pi_k = \int_0^1 q^k dG(q).$$

The role of the mixing distribution in a mixture model is analogous to that of the copula in a latent variable model.

It can be shown that in exchangeable models for large homogeneous groups with similar exposures the tail of the loss distribution is proportional to the tail of the mixing distribution (Frey & McNeil 2001).

©2001 (Alexander McNeil and Rüdiger Frey)

Tail of the mixing distribution – similar default correlation



Tail of the mixing distribution G in three exchangeable Bernoulli mixture models: probit-normal; logit-normal; beta.

©2001 (Alexander McNeil and Rüdiger Frey)

More General Bernoulli Mixture Models

Definition: (Mixture Model with Factor Structure)
$(Y_1,\ldots,Y_m)$ follow a Bernoulli mixture model with p-factor structure if there is a random vector $\Psi = (\Psi_1,\ldots,\Psi_p)$ with $p < m$ and continuous functions $f_i : \mathbb{R}^p \to (0,1)$, such that
1. $Y_i \mid \Psi \sim Be(Q_i)$, $i = 1,\ldots,m$, where

$$Q_i = f_i(\Psi_1,\ldots,\Psi_p) \text{ for all } 1 \le i \le m.$$

2. $(Y_1,\ldots,Y_m)$ are conditionally independent given Ψ

Remark: Poisson mixture models with factor structure can be defined analogously, by making the Poisson rate parameters dependent on Ψ.

Example: CreditRisk$^+$ has this kind of structure.
Moreover, this structure underlies the latent variable models used in practice, as will be seen.

3. Mapping Latent Variable to Mixture Models

It is often possible to transform a latent variable model to obtain an equivalent Bernoulli mixture model with factor structure. This is useful in Monte Carlo simulation, since Bernoulli mixture models are generally easier to simulate than latent variable models.

Example: KMV/Creditmetrics
$\mathbf{X}$ is Gaussian and follows a classical linear p-factor model.

$$X_i = \sum_{j=1}^{p} a_{i,j}\Theta_j + \sigma_i\varepsilon_i$$

for an l-dimensional random vector $\Theta \sim N_p(\mathbf{0},\Omega)$, *independent* standard normally distributed rv's $\varepsilon_1,\ldots,\varepsilon_m$, which are also *independent* of Θ.

CreditMetrics/KMV as a Bernoulli Mixture Model

For the mixing factors take $\Psi = \Theta$.

$$P(Y_i = 1 \mid \Psi) = P(X_i \le D_i \mid \Psi) = P\left(\varepsilon_i \le \left(D_i - \sum_{j=1}^{l} a_{i,j}\Psi_j\right)/\sigma_i\right) = \Phi\left((D_i - \mathbf{a}_i'\Psi)/\sigma_i\right).$$

Clearly $Y_i \mid \Psi \sim \text{Be}(Q_i)$ where $Q_i = \Phi\left((D_i - \mathbf{a}_i'\Psi)/\sigma_i\right)$.

Thus Q_i has a probit-normal distribution.

Moreover, conditional on Ψ, the Y_i are independent.

Mapping Other L.V. Models to Bernoulli Mixture Models

A very similar mapping is often possible when the latent variables follow a multivariate normal mixture model, as in the case of t distributed latent variables or generalised hyperbolic latent variables.

$\mathbf{X}$ has a normal mixture distribution if $X_i = g_i(W) + WZ_i$ where
$W \ge 0$ is independent of $\mathbf{Z}$,
$g_i : (0,\infty) \to \mathbb{R}$,
and $\mathbf{Z}$ is Gaussian vector with $E(\mathbf{Z}) = \mathbf{0}$.

If this underlying Gaussian vector $\mathbf{Z}$ follows a linear factor model as before then it is possible to derive explicitly an equivalent Bernoulli mixture model.

Examples:

1. Student t model: $W = \sqrt{\nu/V}$, $V \sim \chi^2_\nu$ and $g_i(W) = \mu_i$
2. Generalized hyperbolic: $W \sim$ NIG and $g_i(W) = \mu_i + \beta_i W$.

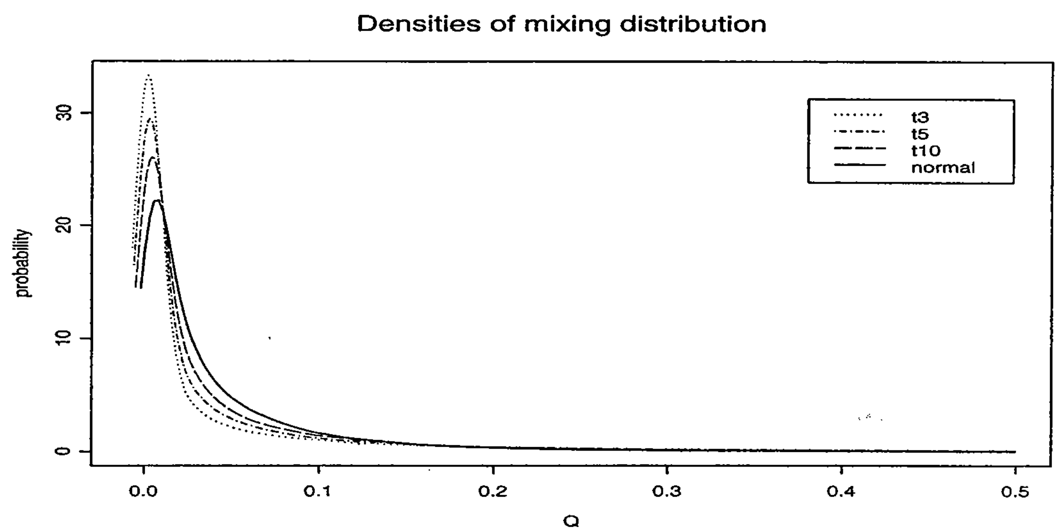
Normal Compared with t using Equivalent Mixture Approach

The profound differences between the Gaussian and t copulas that we observed can now be understood in terms of the differences between the mixture distributions in the equivalent mixture models. Consider two cases (again in exchangeable special case).

Case 1: Asset Correlation held fixed.
Here we observe clear differences between the densities of the equivalent mixing distributions as we vary degrees of freedom. These account for differences in distribution of number of defaults (or total loss distribution.)

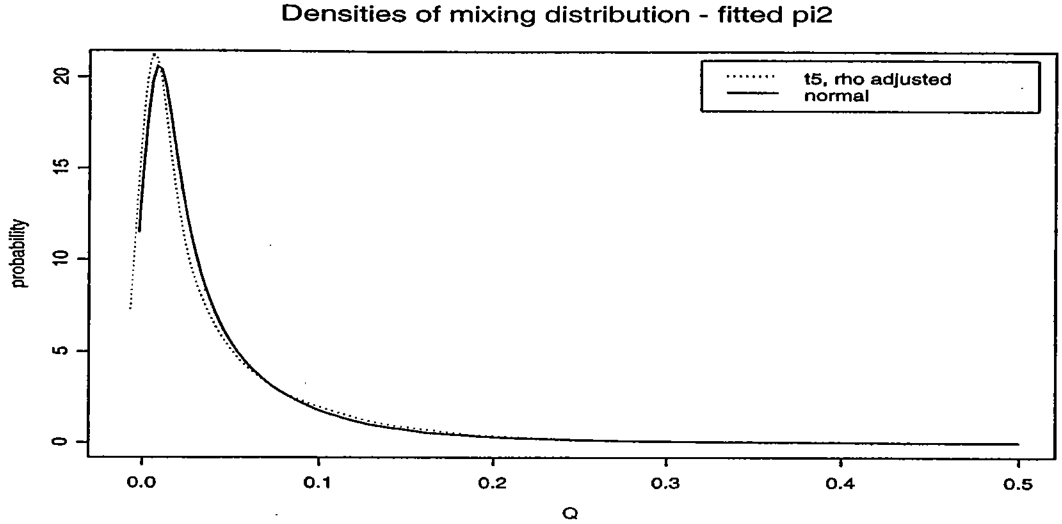
Case 2: Default Correlation held fixed.
Here the differences between the densities are much less obvious. The distributions of the number of defaulting obligors do not differ much at 95th and 99th percentiles. We have to go much further into the tail to see differences.

Densities of mixing distribution – similar asset correlation



Distribution of (Q) for exchangeable Gaussian and t copulas; $\pi = 0.04$ and $\rho = 0.3$.

Densities of mixing distribution – similar default correlation



Distribution of Q for exchangeable Gaussian and t copulas; $\pi = 0.04$ and in the normal model $\rho = 0.3$.

©2001 (Alexander McNeil and Rüdiger Frey)

Estimating Default Probability and Default Correlation

We have seen the importance of π and π_2 (or ρ_Y) in homogeneous groups. How do we estimate these from historical default data for that group?

Suppose our time horizon of interest is one year and we have n years of historical data $\{(m_j, M_j),\ j = 1, \ldots, n\}$, where m_j denotes the number of obligors observed in year j and M_j is the number of these that default.

Assume an exchangeable Bernoulli mixture model in each year period with $Q_1, \ldots, Q_n$ identically distributed. An unbiased and consistent estimator of π_k is

$$\widehat{\pi}_k = \frac{1}{n}\sum_{j=1}^{n} \frac{M_j(M_j-1)\cdots(M_j-k+1)}{m_j(m_j-1)\cdots(m_j-k+1)}, \quad k = 1, 2, 3, \ldots.$$

©2001 (Alexander McNeil and Rüdiger Frey)

4. Implications for pricing basket credit derivatives

Insights on dependence-modelling for loan portfolios have also implications for pricing of basket credit derivatives. Consider portfolio with m obligors (the basket) held by bank A. We are interested in pricing of following stylized default swap:

Second to default swap: Fix horizon T. Bank A receives from counterparty B a fixed payment K at time T if at least two obligors in the basket have defaulted (i.e. had a credit event) until time T; otherwise it receives nothing. At $t = 0$ A pays to B a fixed premium.

Intuition: pricing of this product sensitive to occurrence of joint defaults.

Remark: Real second-to-default swaps are more complicated. The payments depend on identities of defaulted counterparties; moreover, payment due at time of credit event.

©2001 (Alexander McNeil and Rüdiger Frey)

A pricing model

Stylized version of reduced-form model à la Duffie-Singleton or Jarrow-Lando-Turnbull. Our simplifications:

- interest-rate r is deterministic
- default-intensities are rv's instead of processes.

Denote by τ_i the default-time of obligor i in the basket.

Assumption 1: The default-times τ_i, $1 \leq i \leq m$ follow a *mixed exponential distribution*, i.e. there is some p-dimensional random vector Ψ $(p < m)$ such that conditional on Ψ the τ_i are independent exponentially distributed rv's with parameter $\lambda_i(\Psi)$. In particular,

$$P(\tau_i < T|\Psi) = 1 - \exp(-\lambda_i(\Psi)T) \approx \lambda_i(\Psi)T \qquad (1)$$

The default-indicators $Y_i = 1_{\{\tau_i \leq T\}}$ then follow a Bernoulli-mixture model with default-probability as in (1).

©2001 (Alexander McNeil and Rüdiger Frey)

Pricing of credit-derivatives

Following standard-practice we assume that Assumption 1 holds under a pricing-measure Q. Hence for every claim H depending on $\tau_1, \ldots, \tau_m$ the price at $t = 0$ equals

$$P_0 = e^{-rT} E(H(\tau_1, \ldots, \tau_m)).$$

In particular we get for our second-to-default swap

$$P_0 = e^{-rT} Q\left(\sum_{i=1}^{m} Y_i \geq 2\right).$$

Specific model: We choose λ and Ψ so that the one-year default probability corresponds to the default-probability in the one-factor latent variable model with t copula, i.e.

$$\lambda_i = -\ln\left(1 - \Phi\left(\frac{t_\nu^{-1}(\pi)\sqrt{W/\nu} - \sqrt{\rho}\Theta}{\sqrt{1-\rho}}\right)\right), \quad \Theta \sim N(0,1), W \sim \chi^2(\nu)$$

©2001 (Alexander McNeil and Rüdiger Frey)

Simulations: We ran a number of simulations in a homogeneous portfolio with $m = 14$, $T = 1$, and varying values for default probability π and asset correlation ρ.

Portfolio A: $\pi = 0.15\%$ $\rho = 0.38\%$
Portfolio B: $\pi = 0.50\%$ $\rho = 3.80\%$

In the following table we give the ratio $P_0^t / P_0^{\text{normal}}$ of the price of stylized second-to-default swap in in t-model and normal model.

Portfolio	$\nu = 5$	$\nu = 10$	$\nu = 20$
A	11.0	7.3	4.4
B	3.3	2.6	2.0

Choice of the copula has again drastic effect!

Remark (Extension to more realistic model): Model default-times as first jump of a Cox-process where the factor-process governing jump-intensities has stochastic volatility.

©2001 (Alexander McNeil and Rüdiger Frey)

Conclusions

- Extreme risk in latent variable models is driven by the *copula* of the latent variables.
- The assumption of a multivariate normal distribution and a *calibration based on asset correlations* alone may seriously *underestimate* the extreme risk in latent variable models.
- Extreme risk in Bernoulli mixture models with factor structure is driven by the mixing distribution of the factors.
- The two model types may often be mapped into one another. It is particularly useful (Monte Carlo simulation and also for fitting) to represent latent variable models as Bernoulli mixture models.
- Model calibration should use historical default data and not be based solely on assumptions about asset value correlations.
- Statistical fitting issues should now be a research priority.

References

On credit risk modelling:

- Gordy (2000). A comparative anatomy of credit risk models. Journal of Banking and Finance 24.
- Crouhy, Galai & Mark (2000). A comparative analysis of current credit risk models. Journal of Banking and Finance 24.
- Frey & McNeil (2001). Modelling dependent defaults. Preprint at `http://www.math.ethz.ch/~frey`

V. Other Applications: Market Risk and Risk Integration

- Multivariate risk factors - the empirical evidence
- Multivariate time series models
- Multivariate extreme value theory
- Analysing market data - a case study
- Integrating market and credit risk: the role of copulas

©2001 (Alexander McNeil and Rüdiger Frey)

1. Stylized facts of Financial Time Series

A realistic time series model should reflect the *stylized facts* of financial return series :

*Returns not iid but correlation low

*Absolute returns highly correlated and cross-correlated

**Volatility* changes randomly with time

*Returns are *heavy-tailed*

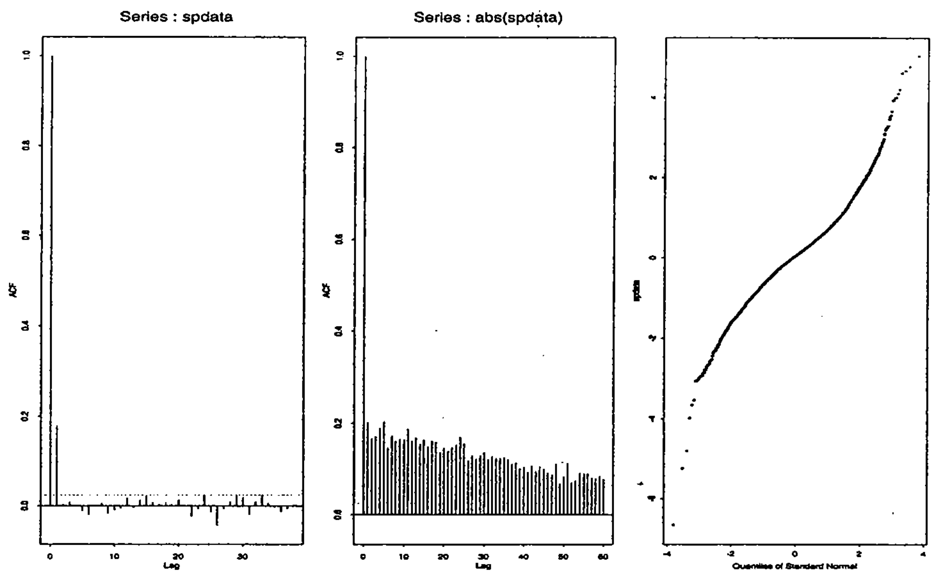
**Extremes* appear in clusters (over time)

*Extreme market moves occur together (across assets)

©2001 (Alexander McNeil and Rüdiger Frey)

Stylized Facts: Correlation, Heavy Tails

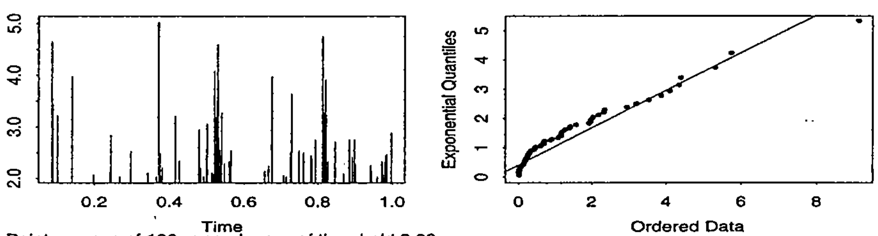
Correlograms of raw S&P data and absolute data, and QQ-plot of raw data.



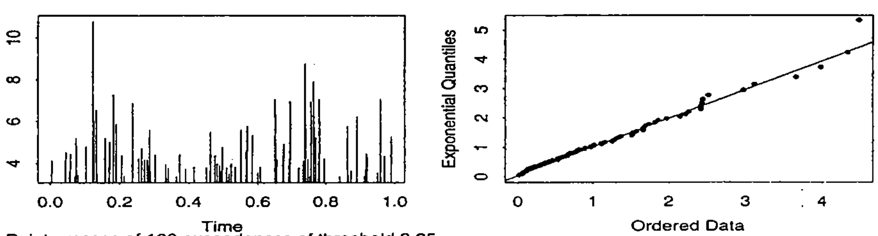
©2001 (Alexander McNeil and Rüdiger Frey)

Stylized Facts: Clustered Extreme Values

Real data show clustered extremes (upper pictures); simulated iid data do not (lower).



Point process of 100 exceedances of threshold 2.03



Point process of 100 exceedances of threshold 3.35

©2001 (Alexander McNeil and Rüdiger Frey)

2 Multivariate Time-Series Models for Market Risk

Goal: Introduction of a few GARCH-type multivariate time-series models which are useful in the context of market risk management; *no* exhaustive overview.

a) Univariate models: a reminder

Basic model structure: Returns follow stationary time series $(X_t, t \in \mathbb{Z})$ with *stochastic volatility*

$$X_t = \mu_t + \sigma_t Z_t, \quad \text{where}$$

(1) μ_t (conditional mean) and σ_t (conditional s.d.) depend on past returns,
(2) Innovations (Z_t) are a (0,1) strict white noise (iid) with df $F_Z(z)$,
(3) X_t is strictly stationary, marginal df $F_X(x)$.

Principal examples: Models from the ARCH and GARCH family.

©2001 (Alexander McNeil and Rüdiger Frey)

Univariate ARCH and GARCH Processes

For simplicity, consider $X_t = \sigma_t Z_t$.

ARCH(p): $\sigma_t^2 = \beta + \sum_{j=1}^{p} \lambda_j X_{t-j}^2$, where $\beta, \lambda_j > 0, \forall j$.

GARCH(p,q): $\sigma_t^2 = \beta + \sum_{j=1}^{p} \lambda_j X_{t-j}^2 + \sum_{k=1}^{q} \delta_k \sigma_{t-k}^2$,
where $\beta, \lambda_j > 0, \forall j$, $\delta_k > 0, \forall k$.

Condition for second-order stationarity: $\sum_{j=1}^{p} \lambda_j + \sum_{k=1}^{q} \delta_k < 1$.

Our main model: GARCH(1,1)

$$\sigma_t^2 \;=\; \alpha_0 + \alpha_1 (X_{t-1} - \mu_{t-1})^2 + \beta \sigma_{t-1}^2,$$

with α_0, α_1, $\beta > 0$ and $|\lambda| < 1$.

Many other GARCH-models exist. Estimation is typically done with (quasi)MLE.

©2001 (Alexander McNeil and Rüdiger Frey)

b) Multivariate GARCH-models

Basic model structure: Observations $(\mathbf{X}_t)_{t\in\mathbb{Z}}$, where $\mathbf{X}_t = (X_t^1, \ldots, X_t^d)$ is d-dim random vector. Dynamics of $\mathbf{X}_t$ are of the form $\mathbf{X}_t = \mu_t + \Sigma_t^{1/2}\mathbf{Z}_t$, where

- μ_t is conditional mean-vector of $\mathbf{X}_t$ given past observations,
- $\Sigma_t = \Sigma_t^{1/2}\left(\Sigma_t^{1/2}\right)^t$ is conditional covariance-matrix of $\mathbf{X}_t$,
- and where $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$ is an iid-sequence with $\text{cov}(\mathbf{Z}_t) = \text{id}$.
- Typically a multivariate normal or t-distribution is used for $\mathbf{Z}$.

Conditional correlation: Denote by σ_t^i the conditional s.d. of X_t^i. Conditional correlation matrix is given by $(C_t)_{i,j} = (\Sigma_t)_{i,j}\big/(\sigma_t^i\sigma_t^j)$. Note that

$$\Sigma_t = D_t C_t D_t, \text{ where } D_t = \text{diag}(\sigma_t^1, \ldots, \sigma_t^d)$$

The constant conditional correlation model

Assume $\mu_t = 0$. In a general multivariate GARCH-model D_t and C_t depend on its own past values and on all past values of $\mathbf{X}_{t-1} \otimes \mathbf{X}_{t-1}$ (productwise multiplication). For practical application a more parsimonious version is needed. A useful model is

Constant conditional correlation model (CCC-model): In this model, proposed by Bollerslev (1990) it is assumed that

- Conditional correlation is constant over time, $C_t \equiv C$ for all t
- Conditional s.d. is given by standard univariate GARCH $(\sigma_t^i)^2 = \alpha_0^i + \alpha_1^i\left(X_{t-1}^i\right)^2 + \beta(\sigma_{t-1}^i)^2$,
- $\mathbf{Z}$ is normal or t-distributed

CCC-model - Estimation

Estimation is usually done via MLE. Often the matrix C is taken as (robust version of) the sample correlation matrix $\widehat{C}$; alternatively $\widehat{C}$ is taken as starting value for optimization-procedure in MLE-estimation.

An alternative: (Cross-sectional returns via historical simulation)

In CCC-model we have $\mathbf{X}_t = D_t \mathbf{Y}_t$ where the sequence $(\mathbf{Y}_t)_{t\in\mathbb{Z}}$ is iid and $\mathbf{Y} \sim N(0, C)$ resp. $\mathbf{Y} \sim t_\nu(0, C)$. Alternatively we could use the *empirical distribution* of standardized residuals as model for distribution of $\mathbf{Y}_t$.

As in the CCC-model standardization is done by fitting one-dimensional GARCH-models to the components and by defining $Y_t^i := X_t^i/\sigma_t^i$. The standardized residuals can be stored and used for simulations, e.g. in the context of pricing *basket derivatives*.

3. Extreme Value Theory

Univariate EVT

In classical univariate EVT there are two related modelling approaches:

1. Modelling block maxima (e.g. yearly maxima of daily data).
2. Modelling threshold exceedances - the POT method

Multivariate EVT

In multivariate EVT these modelling approaches are extended:

1. Modelling *componentwise* block maxima of vectors.
2. Modelling *joint* exceedances of several thresholds.

We concentrate on the latter, whose utility is more obvious. We concentrate on building theoretically supported models for tails of distributions.

Univariate EVT

We use *parametric methods* based on *GPD* (generalized Pareto distribution). Alternative: semi-parametric methods based on *Hill estimator* of tail index.

Generalized Pareto Distribution

The GPD is a two parameter distribution with df

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi} & \xi \neq 0, \\ 1 - \exp(-x/\beta) & \xi = 0, \end{cases}$$

where $\beta > 0$, and the support is $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ when $\xi < 0$.

Moments: For $\xi > 0$ distribution is heavy tailed. $E(X^k)$ does not exist for $k \geq 1/\xi$.

A Key Result in EVT

Define the distribution function of excesses losses over a high threshold u to be

$$\begin{aligned} F_u(x) &= P\{X - u \leq x \mid X > u\} \\ &= \frac{F(x+u) - F(u)}{1 - F(u)}, \end{aligned}$$

for $0 \leq x < x_0 - u$ where $x_0 \leq \infty$ is the right endpoint of F.

Theorem: for a *large class* of underlying distributions we can find a function $\beta(u)$ such that

$$\lim_{u \to x_0} \sup_{0 \leq x < x_0 - u} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0.$$

This class consists of the distributions in the *maximum domain of attraction* of an *extreme value distribution* and includes *all common* continuous distributions. The GPD is thus a *natural approximation* to the unknown excess distribution above sufficiently high thresholds.

The POT Method - Smith's Estimator (1987)

We have data $X_1, \ldots, X_n$ from F.

Fix u, a high threshold, and count the random number of exceedances N_u. The GPD is fitted to the excesses above u by maximum likelihood to obtain estimates $\hat{\xi}$ and $\hat{\beta}$.

By noting that for $x > u$ the tail can be written

$$\begin{aligned} \bar{F}(x) &= P\{X > u\} P\{X > x \mid X > u\} \\ &= \bar{F}(u)\bar{F}_u(x-u), \end{aligned}$$

we arrive at the *tail estimator*

$$\hat{\bar{F}}(x) = 1 - \frac{N_u}{n}\left(1 + \hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}};$$

this estimator is only valid for $x > u$. Properties (consistency, asymptotic normality) are established in the iid case.

©2001 (Alexander McNeil and Rüdiger Frey)

Estimating Measures of Tail Risk

Assume that $\mathrm{VaR}_q(X) > u$, so that risk measures are beyond EVT threshold.

* Quantile (VaR) Estimator

$$\widehat{\mathrm{VaR}}_q(X) = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{1-q}{N_u/n}\right)^{-\hat{\xi}} - 1\right).$$

Obtained by inverting tail estimator.
* Expected Shortfall

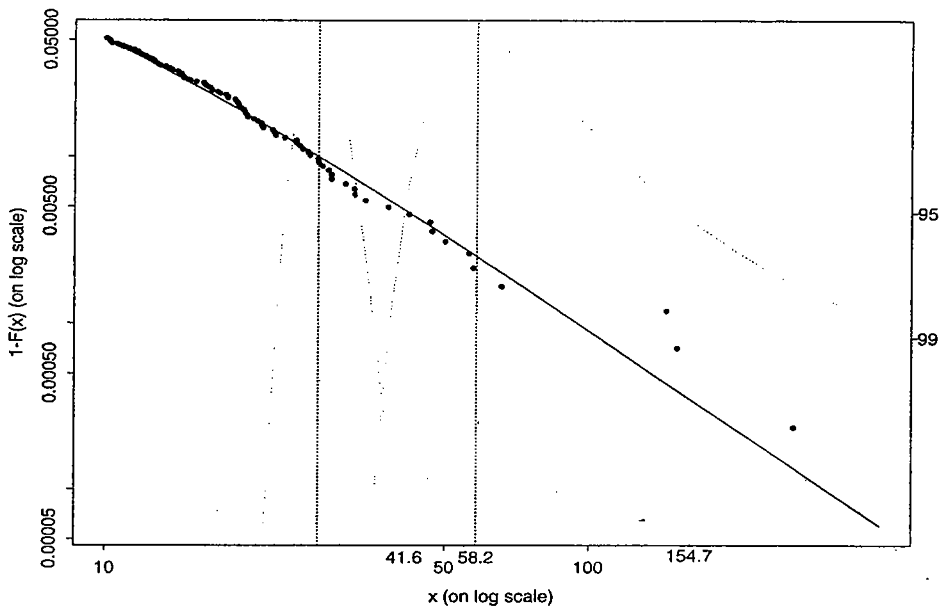
$$\widehat{\mathrm{ES}}_q(X) = \frac{\widehat{\mathrm{VaR}}_q(X)}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{1-\hat{\xi}}.$$

Obtained by simple calculation due to linear form of mean excess function for GPD.

Asymmetric *confidence intervals* for these estimates can be constructed using profile likelihood method.

©2001 (Alexander McNeil and Rüdiger Frey)

Univariate EVT Analysis: 99% VaR and Shortfall Estimates



Multivariate POT Analysis

Now consider iid random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ and d thresholds $u_1, \ldots, u_d$. We seek an estimate for the joint distribution $F(x_1, \ldots, x_d)$ which is valid in the area $\{\mathbf{x} : x_1 > u_1, \ldots, x_d > u_d\}$.

Suitable margins for our model have already been suggested by univariate EVT. We now require a copula to complete the specification of our model.

Theory suggests we should choose a so-called extreme value copula having property: $C(u_1^t, \ldots, u_d^t) = C^t(u_1, \ldots, u_d)$.

There are many such copulas. An example is the Gumbel copula.

Example: Bivariate POT Model

We model the margins with the POT tail estimators

$$\widehat{F}_i(x) = 1 - \frac{N_{u_i}}{n}\left(1 + \hat{\xi}_i \frac{x - u_i}{\hat{\beta}_i}\right)^{-1/\hat{\xi}_i}, \; x > u_i, \; i = 1, 2.$$

We model the dependence with the Gumbel copula C_{β}^{Gu}, for some estimated β.

We have the joint model

$$\widehat{F}(x_1, x_2) = C_{\beta}^{\text{Gu}}\left(\widehat{F}_1(x), \widehat{F}_2(x)\right), x_1 > u_1, \; x_2 > u_2.$$

Statistics. We may estimate the marginal and copula parameters in two stages (as before) or all together.

Inference. Can calculate joint extreme probabilities or *spillover* probabilities: $P(X_1 > k_1, X_2 > k_2)$ or $P(X_2 > k_2 \mid X_1 > k_1)$ for $k_1 > u_1$ and $k_2 > u_2$,

4. Towards Risk Integration: pragmatic solutions

It is often inevitable that risk measurement in a financial institution is decentralised. Individual units or departments know their own risks and have their own data with which they calculate their own P&L.

These may be considered to be the marginal distributions in a multivariate risk model representing the whole institution. The missing element is the copula expressing the dependence between risks in different units.

Example (with some basis in reality)
Calculation of market and credit risk P&L's is decentralised. Both departments are proud of their own risk assessments. But management needs them integrated.

Copula-Based Solution (as implemented by financial institution X)

1. Build an economic factor model with one set of common factors (such as market indexes in various sectors and countries) which are considered to impact both market and credit risks.

2. Model underlying factors using distributions that are heavy-tailed and tail dependent to take into account extreme risks!

3. Simulate market and credit losses using the factor model and use these simulate data to estimate a suitable bivariate copula.

4. Combine estimated copula with original, locally estimated P&Ls to produce joint model.

References

On multivariate time series:

- Bollerslev (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. Review of Economics and Statistics.

On multivariate extreme value theory:

- Joe (1997). Multivariate Models and Dependence Concepts. Chapman & Hall, London.

On multivariate extremes and multivariate time series:

- Starica (1999). Multivariate extremes for models with constant conditional correlations. Journal of Empirical Finance, 6.