Structured copula modelling of insurance liabilities

Piet de Jong

Department of Actuarial Studies,
Macquarie University, NSW 2109, Australia.
Email: piet.dejong@mq.edu.au

Abstract

This article employs copulas to model total claims losses arising from different lines of insurance. There are two contributions. First a flexible method is developed to generate copulas. Second total losses are considered given the marginal distributions are Weibull.

Keywords: Copulas

1 Introduction

This article develops and discusses a practical, structural way of generating copulas. By structural we mean the copula is generated from a generating mechanism typical of statistical modeling. This contrasts with the usual way of specifying copulas directly via functional forms as discussed for example in joe1997mma or nelsen2006ic.

Figure 1 displays simulations from a copula constructed using the proposed copula generating method. Each panel in Figure 1 is generated by firstly simulating for $t = 1, \ldots, N$ from the bivariate distribution

$$s_{1t} = \alpha_t + \psi_1 \epsilon_{1t} \ , \qquad s_{2t} = \alpha_t + \psi_2 \epsilon_t \ , \qquad \alpha_t = u_t^{-\lambda} \ , \qquad \epsilon_t \sim N(0,1) \ . \quad (1)$$

where u_t is a uniform random variable and $0 < \lambda < 1$. Given simulated values s_{1t} and s_{2t}, the corresponding percentiles $\mathrm{P}(s_t)$ are calculated relative to each marginal distribution associated with s_t. Since marginals are not

specified directly, the percentiles are estimated by the percentile ranks $\check{s}_t$: their ranks in the simulated sample divided by the sampling effort N.

The specification $\alpha_t = u_t^{-\lambda}$ where u_t is uniform covers a broad range of setups. In particular for $\lambda > 0$ and $c > 0$, the percentiles of $s_t = 1'\alpha_t + \psi \cdot \epsilon_t$ under this setup coincide with those if

$$s_t = 1c\left(\frac{u_t^{-\lambda} - 1}{\lambda}\right) + \frac{c}{\lambda}\psi \cdot \epsilon \ .$$

Letting $\lambda \to 0$, then the first fraction on the right converges to $\ln u_t$ while $c\psi/\lambda$ under appropriate choice of $\psi \to 0$ converges to a constant vector. Hence the above setup contains the setup $\alpha_t = -c \ln u_t$ which implies α_t is exponential with mean $1/c$. Thus for $\lambda \approx 0$ the above setup corresponds to an exponential common factor.

2 Relationship to Burr distribution and its variants

In this section we consider the distribution of $c(u_t^{-\lambda}-1)$ and its generalization $c(u_t^{-\lambda}-1)^\gamma$. Here u_t is uniform while $c > 0$ and $\gamma > 0$. The former is a special case of the latter by setting $\gamma = 1$.

Note that insofar as percentiles are concerned,

$$\mathrm{P}\left\{1c(u_t^{-\lambda} - 1) + \psi \cdot \epsilon_t\right\} = \mathrm{P}\left(1u_t^{-\lambda} + \frac{1}{c}\psi \cdot \epsilon_t\right)$$

What is the formal definition of P. It is $\mathrm{P}(x) = F(x)$ where x is the random variable of which P is taken. Note that x here has a dual role: indicating both the random variable (and hence its distribution) and the the value of the random variable at which the distribution is being evaluated.

Note for any monotonically increasing transformation f, $\mathrm{P}\{f(u_t)\} = \mathrm{P}(u_t) = u_t$. The last equality holds because u_t is uniform. Hence

$$\mathrm{P}\{1c(u_t^{-\lambda} - 1)^\gamma\} = \mathrm{P}(1u_t) = 1u_t \ .$$

This drives home the point that the only role of c, λ and γ is to modify the role of the error ϵ_t.

- The distribution of $cu_t^{-\lambda} - 1$ has a hazard rate

$$\frac{1}{\lambda(x + c)} \ .$$

  Lomax used this kind of hazard rate to fit business failure data and the corresponding distribution is sometimes called the Lomax distribution.

- Suppose conditional on β, x has the exponential distribution $1 - \mathrm{e}^{-\beta x}$ with density $\beta \mathrm{e}^{-\beta x}$ and β has the Gamma distribution. Then the distribution of x is the same as that of $c(u_t^{-\lambda} - 1)$. Thus we have an exponential mixed by a gamma gives the Burr distribution. This means the the setup $u_t^{-\lambda}$ can be viewed as arising form gamma distributed exponential random variables.

- What is Pareto about $u_t^{-\lambda}$?

- A generalization is the Burr distribution: $\alpha_t = (u_t^{-\lambda} - 1)^{\gamma}$ equivalent to the very popular "q–exponential" distribution extensively used in statistical physics. Here $\lambda > 0$ and $\gamma > 0$. The case $\gamma = 1$ is the Pareto setup discussed before. The case $\alpha_t = (-\ln u_t)^{\gamma}$ corresponds to the Weibull distribution.

3 Time series modeling of a copula

Consider the setup

$$s_t = 1u_t^{-\lambda} + \sigma_t \psi \cdot \epsilon_t \ ,$$

where $\sigma_t > 0$. This generates the same copula as $s_t = 1\sigma_t^{-1} u_t^{-\lambda} + \psi \cdot \epsilon_t$. If σ_t is large then the independent Gaussian terms predominate and there is little dependence. If σ_t is small the component $u_t^{-\lambda}$ predominates and there is increasing tail dependence. Hence σ_t is a measure of tail dependence.

Now write

$$\sigma_t = \mathrm{e}^{h_t/2} \ , \qquad h_{t+1} = a h_t + \eta_t \ ,$$

where $|a| < 1$ and η_t is zero mean noise with constant variance. Thus $h_t = \ln(\sigma_t^2)$ is modeled as a zero mean autoregressive process of order 1. Increasing values of h_t translate to increasing dependence. The above model is termed the stochastic copula model which is obviously closely related to stochastic volatility models.

Leverage effects are modeled by correlating η_t with ϵ_t in which case the direction of independent components in ϵ_t influence future tail dependence. For example if the ϵ_t are positively correlated with η_t and all the independent components are below their mean indicating the percentiles $\check{s}_t$ are relatively low, then h_t is likely to increase implying a fall in σ_t. In turn this means increasing future dependence.

Note h_t is stationary and hence $\mathrm{E}(h_t) = 0$ and variance $\sigma_\eta^2/(1-a^2)$. Hence

$$\mathrm{E}(\mathrm{e}^{-h_t/2}) = \mathrm{e}$$

The first term on the right defines the Weibull.

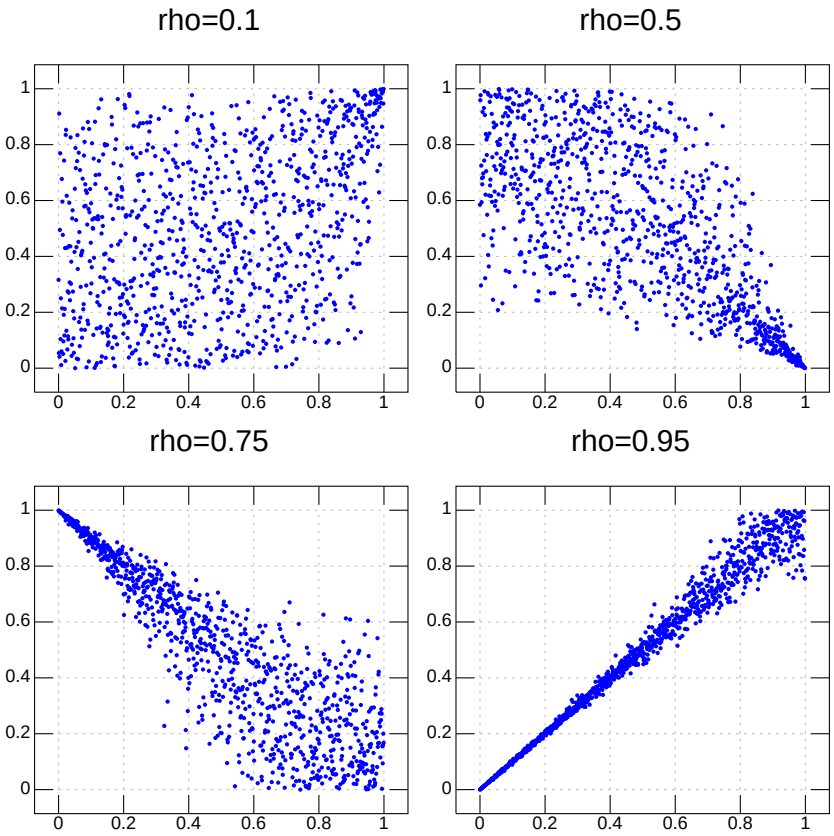

Figure 1: Simulated percentile ranks $(\check{s}_{it}, \check{s}_{2t})$ from model (1).

In Figure 1, if $\rho \approx 0$ then $\ln(\rho)$ is very negative and s_{1t} and s_{2t} are virtually unrelated since s_{2t} is dominated by $\ln(\rho)\epsilon_t$. If $\rho \approx 1$ then $\ln(\rho) \approx 0$ and s_{1t} and s_{2t} are highly dependent. The panels in Figure 1 suggest increasing upper tail dependence in that the conditional probability $P(\check{s}_{2t} > u | \check{s}_{1t} > u)$ increases as $u \to 1$.

The chi–squared specification for α_t can be changed to other extreme value or heavy tailed distributions. For example the log–normal is an alternative. Figure 2 illustrates the copula when $\alpha_t \sim t_1$, in (1), a "t" random variable with 1 degree of freedom. In this case tail dependence appears to increases in both tails. Thus when α_t is near zero, the effect of ϵ_t is to mask the dependence while if $|\alpha_t|$ is large the dependence increases: the perturbing effect of ϵ_t in these cases is negligible.

The copula generating method (1) generalizes the Gaussian copula [McNeil, Frey, Embrechts, e
If $\alpha_t \sim N(0,1)$ then s_{1t} and s_{2t} are jointly normal with correlation $1/\sqrt{1+(\ln \rho)^2}$.
Thus in this case the relationship between the percentiles is induced with a

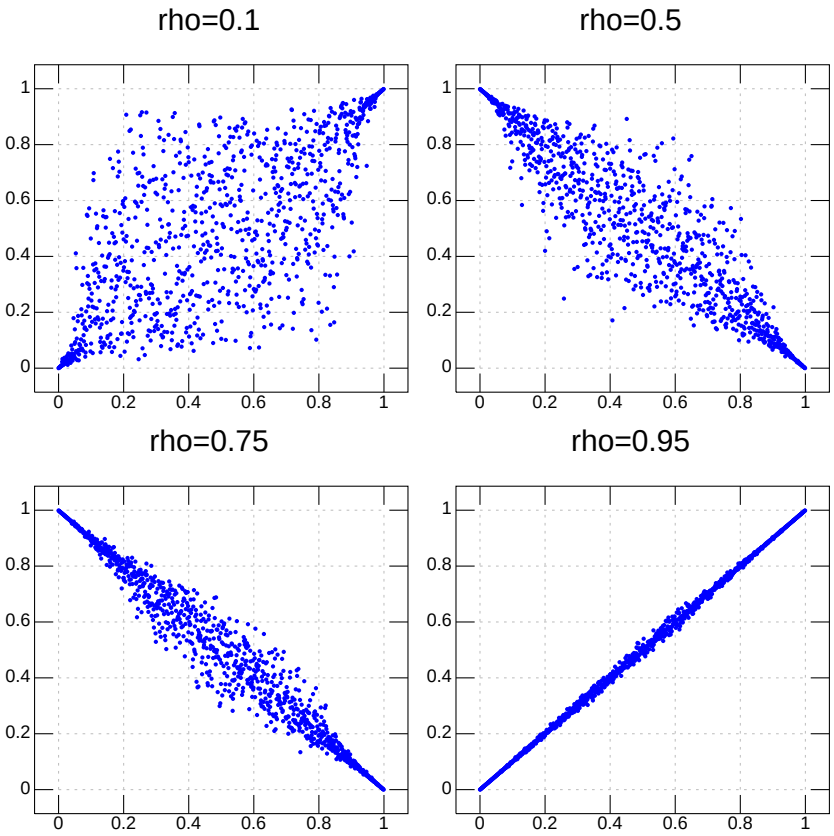

Figure 2: Simulated percentile ranks from model (1) if $\alpha_t \sim t_1$.

Gaussian copula. The Gaussian copula combined with Gaussian marginal distributions for the actual observed variables yields the usual factor model often employed in the study of correlated data [Lawley and MaxwellLawley and Maxwell1971].

4 The general form of the model

Setup (1) can be written in the vector form

$$\begin{pmatrix} s_{1t} \\ s_{2t} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \alpha_t + \begin{pmatrix} 0 & 0 \\ 0 & \ln \rho \end{pmatrix} \epsilon_t \ , \tag{2}$$

This suggests the generalization

$$\begin{pmatrix} s_{1t} \\ s_{2t} \\ \vdots \\ s_{pt} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \alpha_t + \begin{pmatrix} \ln \rho_1 & 0 & \cdots & 0 \\ 0 & \ln \rho_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \ln \rho_p \end{pmatrix} \epsilon_t \ , \qquad \epsilon_t \sim N(0, I) \ , \tag{3}$$

where the vector random variable α_t has an appropriate long tail distribution, and $\rho_1 = 1$.

The general form of the proposed model is

$$y_t = F^-(\check{s}_t) \ , \qquad s_t = Z\alpha_t + \psi \cdot \epsilon_t \ , \qquad \epsilon_t \sim N(0, I) \ . \tag{4}$$

Here $\check{s}_t$ is the vector of percentiles of each component of s_t. Further α_t is a vector of independent random variables with a heavy tailed distribution. Further ψ is a vector with entries $\ln(\rho_j)$ where $\rho_1 = 1$, the $\cdot$ notation indicates componentwise product, and Z is a matrix of factor loadings. The vector function G is the vector of marginal distributions determined the joint distribution of s_t. In particular component i of $G(s_t)$ is $G_i(s_{it})$ where G_i is the marginal distribution of s_{it}, component i of s_t. The vector of functions F is similarly defined in terms of the marginal distributions of y_t, and F^- denotes the vector of inverse mapping.

Analogous to state space models in time series, the first equation is called the measurement equation indicating how the measurements y_t are related to the state s_t. The second equation is called the state equation.

α_t is the "state" vector, driving the tail dependence or tail correlation in the system. Thinking of a one–dimensional state, when α_t is large

The marginal distributions in G are easy to describe and simulate from but functional forms for the specific cases of interest are usually not available. For example if α_t contains chi–squared random variables and ϵ_t is normal then the marginal distributions G correspond to that of linear combinations of chi–squareds and normals.

When functional forms for the components of G are not available it is convenient to employ the nonparametric estimate

$$\check{G}(s_t) \equiv \check{s}_t \equiv \frac{1}{N} \ \varrho(s_t) \ ,$$

where the s_r, $r = 1, \ldots, N$ are simulations from the second equation in (4) and $\varrho(s_t)$ denotes the vector of ranks of each component of s_t.

The vector of marginal distributions F may be functionally specified, up to unknown parameters, in which case the available data y_t, $t = 1, \ldots, n$ is used to estimate the unknown parameters leading to the estimate $\hat{F}$. Alternatively

$$\check{F}(y_t) \equiv \check{y}_t \equiv \frac{1}{n} \ \varrho(y_t) \ .$$

If n is small this estimate yields worthless tail probability estimates of either 0 or 1 arguing for a functional estimate $\hat{F}$ of F. Suggestions are in the next section.

5 Modeling the marginal distributions

This section discusses the modeling of the marginals in F. It is possible to model each marginal by a different form. However in many applications differences will be confined to differences in the parameters, keeping the functional form fixed.

5.1 Normal or log–normal marginals

Suppose $\Phi^-(\check{s}_t)$ is the standard inverse normal evaluated at each of the components of $\check{s}_t$ and let μ and σ be vectors of means and standard deviations, respectively. Then the completed model is

$$y_t = \mu + \sigma \cdot \Phi^-(\check{s}_t) \ , \qquad s_t = Z\alpha_t + \psi \cdot \epsilon_t \ , \qquad \epsilon_t \sim N(0, I) \ .$$

Note that if α_t is normal then $\check{s}_t = \Phi_*(s_t)$ where Φ_* denotes a vector of marginal normals with means and variances as induced by the state equation. Hence the net effect of both measurement and state equation is to induce a jointly normal y_t with correlation matrix derived from the state equation. Hence this setup with α_t normal, returns to the classical factor analysis setup [Lawley and MaxwellLawley and Maxwell1971]

With the log–normal model y_t in the measurement equation is replaced by $\ln y_t$ in which case μ and σ refer to the mean and variance of $\ln y_t$.

Figure 3 displays simulations from (1) if $\alpha_t \sim \chi^2_1$ with F standard normal. Notice for all values of ρ, there are many more "+2" standard deviation events compared to "-2" standard deviation events – that is events that lie in the upper right hand square of each panel as opposed to the lower left panel. The bottom right panel shows moderate correlation for all occurrences below the means 0, with almost perfect correlation for above mean events.

6 Pareto based copulas

The Pareto distribution and its inverse are:

$$u = 1 - \left(\frac{\beta}{\beta + x}\right)^{1/\lambda} \ , \qquad x = \beta \left\{\frac{1}{(1-u)^\lambda} - 1\right\} \ .$$

with $x > 0$ and $0 < u < 1$ and the parameters $\beta > 0$ and $0 < \lambda < 1$. The distribution is used to model extreme events. Here it is used to model tail dependence and in particular extreme tail dependence. That is dependence

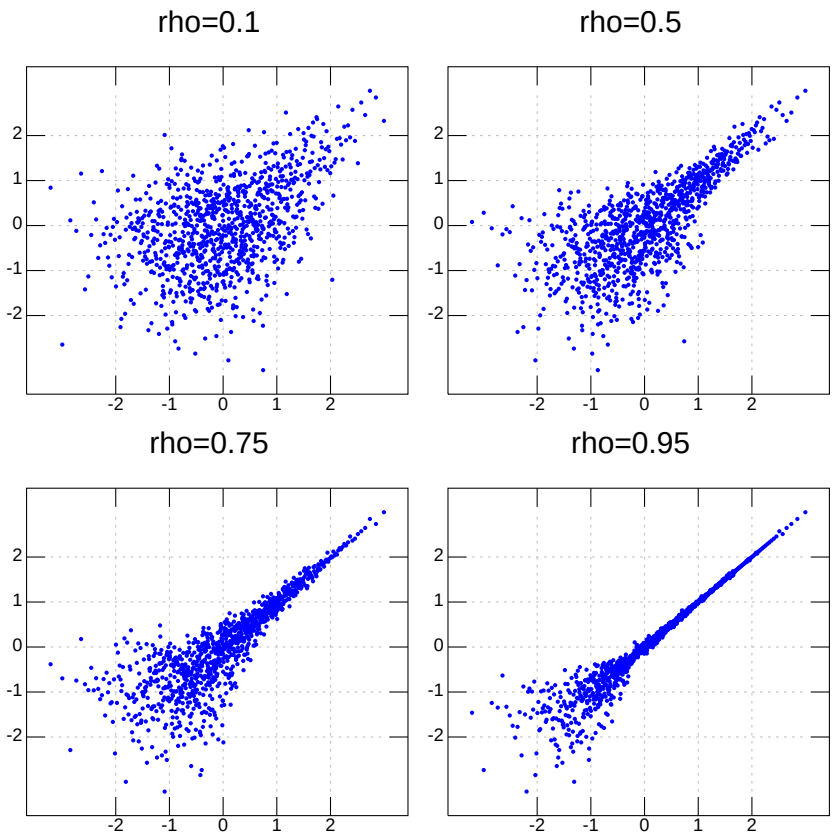

Figure 3: Simulated values from (1) if $\alpha_t \sim \chi^2_1$ with F normal .

that changes as one goes further into the tails of the marginal distributions. The mean and variance are:

$$\frac{\beta}{1-\lambda} \ ,$$

The parameter β is a scale parameter.

The ratio of exceedence probabilities at two points x and ax, $a > 1$ are

$$\left(\frac{\beta + ax}{\beta + x}\right)^{1/\lambda} \approx a^{1/\lambda} \ ,$$

provided x is large in relation to β. A value of λ near 1 indicates a very fat tail.

The inverse distribution indicates a way to generate Pareto variable: First, generate a uniform random u and use in the second formula above to derive the a simulated value x from the Pareto. Since both u and $1-u$

are uniform then $x = \beta(u^{-\lambda} - 1)$ is Pareto. Further the percentiles of and $\beta(u^{-\lambda}-1)+\epsilon$ coincide with those of $u^{-\lambda}+\beta^{-1}\epsilon$ provided $\beta > 0$. This suggests the following copula based generating model:

$$y_t = F^-(\check{s}_t) \ , \qquad s_t = 1u_t^{-\lambda} + \psi \cdot \epsilon_t \ , \qquad \epsilon_t \sim N(0, I) \ ,$$

where ψ is a vector.

Note

$$\mathrm{cov}(s_t) = 1\mathrm{cov}(u_t^{-\lambda})1' + \mathrm{diag}(\psi^2) \ ,$$

This implies the correlation between component i and j of s_t is

$$\left\{\left(1 + \frac{\psi_i^2}{v}\right)\left(1 + \frac{\psi_j^2}{v}\right)\right\}^{-1/2} \approx \ \mathrm{e}^{-(\psi_i^2+\psi_j^2)/2v} \ ,$$

provided $(\psi_i^2 + \psi_j^2)/v$ is small and where, provided $2\lambda < 1$,

$$v \equiv \mathrm{cov}(u_t^{-\lambda}) = \frac{\lambda^2}{(1-\lambda)^2(1-2\lambda)} \ .$$

Thus correlation approaches 1 if $\psi_1^2 + \psi_2^2$ approaches 0, or λ approaches $1/2$.

In the above model all dependence arises from the common factor u_t. The parameter λ controls the severity of tail dependence. When λ is near 1, tail dependence in the sense of dependence between percentiles is huge.

In this model all dependence arises from the common factor u_t. The components of ψ control the overall importance of the common factor and hence control the degree of dependence.

The components of ψ control noise–to–signal ratio. If ψ is large then the Gaussian noise

Figure **??** illustrates the copula based on the Pareto.

If the uniform random variable u_t is replaced by a standard normal z_t with $\lambda = -$ yields a Gaussian copula. If $s_t = 1z_t^{-\lambda} + \psi \cdot \epsilon_t$ with $\lambda = -2$ we have the chi–squared setup of before. Further if u_t is replaced by $\ln u_t$ where u_t is uniform then we have an exponential type setup.

Since $0 < u_t < 1$ and $0 < \lambda < 1$ it follows $1 < u_t^{-\lambda} < \infty$. Hence for small s_t the dependence structure is that of ordinary normal correlation. However large s_t are large because $1/u_t$ is large and if u_t is large then there is high dependence between the components of s_t.

7 Specifying the marginals

Since we are looking at $F(\check{s}_t)$ the question is how to do this nicely. One way is to use Pareto marginals which implies

$$F^-(\check{s}_t) = \beta \cdot \check{s}_t^{-\gamma} - \beta \ ,$$

where all operations are interpreted in a componentwise vector sense. Hence the total liability is

$$1'(\beta \cdot \check{s}_t^{\gamma}) - 1'\beta \ .$$

7.1 Pareto marginals

The inverse of the Pareto distribution G is

$$G^{-}(u) = \beta \left(\frac{1}{\sqrt[\gamma]{1-u_t}} - 1 \right), \qquad 0 < u < 1 \ , \tag{5}$$

where $\gamma > 1$ and $\beta > 0$ are parameters. The mean of the distribution is $\beta/(\gamma - 1)$. This suggests the following structural copula model:

$$y_t = F^{-}(\check{s}_t) \ , \qquad s_t = \beta u_t^{\delta} + \epsilon_t \ , \tag{6}$$

where u_t is uniform, β is a vector of weights, and $\epsilon_t \sim N(0, I)$. In this specification $1 - u_t$ and -1 in (7) have been changed to u_t and 0, respectively since neither of these changes effect percentile ranks. Also the parametrization using γ in (7) is changed to $\delta = -1/\gamma$. This form is closely related to the Box–Cox transform with $\delta = 1$ corresponding to the identity transform.

Consider this model

$$y_t = F^{-}(\check{s}_t) \ , \qquad s_t = 1\frac{u_t^{\lambda} - 1}{\lambda} + \psi \cdot \epsilon_t \ , \qquad \epsilon_t \sim N(0, I) \ .$$

where u_t is a scalar random variable, either the uniform or standard normal. Specialized choices for λ lead to specialized setups.

-
-
-

The Pareto distribution and its inverse used here is parametrized

$$G(x) = 1 - \left(\frac{\beta}{x + \beta} \right)^{\gamma} \ , \qquad G^{-}(u) = \beta \left(\frac{1}{\sqrt[\gamma]{1-u}} - 1 \right) \ ,$$

An alternative specification is based on the logit transform and states

$$s_t = 1 \ln \frac{u_t}{1 - u_t} + \psi \cdot \epsilon_t \ .$$

Combining this with the normal ϵ_t implies that the correlational properties of s_t given $s_t < 1$ are Gaussian. The probability $s_t < 1$ is controlled through the size of ψ. To summarize,

If all the components of ψ approach zero, the components of s_t will be highly dependent. If ψ is moderate, the effect of ϵ_t is dominant except if

The utility of writing things in these ways is the following. As $\lambda \to 0$, or equivalently $\gamma \to \infty$, $(u^\lambda - 1)/\lambda \to \ln u$.

which leads to the model When adding a scaled standard normal and computing the percentile of the resulting random variable is equivalent to computing the percentile of $u^{-/\gamma}$

The inverse of the Pareto distribution is

$$F^-(u) = \beta \left\{ (1-u)^{-1/\gamma} - 1 \right\} , \qquad 0 < u < 1 ,$$

where $\gamma > 1$ and $\beta > 0$ are parameters. The mean of the distribution is $\beta/(\gamma - 1)$. This suggests the following copula model:

$$y_t = F^-(\check{s}_t) , \qquad s_t = \beta\{(1-u_t)^{-1/\gamma} - 1\} + \epsilon_t , \tag{7}$$

where u_t is a uniform random variable, β is a vector of weights, and $\epsilon_t \sim N(0, I)$.

It is worthwhile to consider this model in the two dimensional case. In this case there is no loss in generality in assuming

$$s_{1t} = u_t , \qquad s_{2t} = \beta u_t^\gamma + \epsilon_t .$$

As $\beta \to \infty$ swamp any effect due to the normal error ϵ_t. Note that $\check{s}_{1t} = [u_t]$. When $\beta = 1$

and ϕ is a scalar function. Note that multiplying the right hand side of the state equation by a constant does not alter the the percentile ranks $\check{s}_t$ hence any such constants do not effect the model. On the other hand the vector of ψ does effect the model. Large ψ indicate the components of s_t are unrelated and hence the implied copula is the independence copula. Small ψ on the other hand indicate each component in s_t is highly related to all the other components.

Initially suppose u_t is scalar. Two cases are of particular interest:

- Suppose $\phi(u_t) = \ln(u_t)$. Then $-\phi(u_t)$ has the exponential distribution with mean 1. Note that raising u_t to any power is equivalent to multiplying $\phi(u_t)$ by the power, and hence, as discussed above, does not alter percentiles and hence the structure of the model. Obviously this is closely related to the chi–squared distribution with 1 degree of freedom.

- Suppose $\phi(u_t) = u_t^{\gamma}$. This setup is closely related to a Pareto distribution specification for $\phi(u_t)$. In particular if $\gamma \to 0$

- A specification that covers both of the above prescriptions is

$$\phi(u_t) = \frac{u_t^{\lambda} - 1}{\lambda} \ .$$

  As $\lambda \to 0$, $\phi(u_t) \to \ln u_t$, the first prescription. For $\lambda \neq 0$ the scale adjustment λ can be absorbed into ψ in that percentiles are unaffected. Similarly the location adjustment $-1/\lambda$ has no affect insofar as percentiles are concerned. Note that the above specification is the Box–Cox transform of a uniform random variable.

What happens if u_t in the above is replaced by a standard normal random variable. Then if ϕ is an identity transform and $\phi = 1$ then s_t is multivariate normal with covariance matrix $11' + \text{diag}(\psi^2)$. In the two dimensional case this yields an ordinary Gaussian copula.

Consider for example Then $-\delta \ln u_t$ has an exponential distribution with mean δ.

If where β is a vector then each component of $\phi(u_t)$ is a scaled version of the Pareto distribution with

onsider a small interval centered on $0 < u < 1$. Then a first order Taylor expansion yields

$$\beta u_t^{\delta} + \epsilon_t \approx \beta u^{\delta} + \beta(u_t - u)\delta u^{\delta - 1} + \epsilon_t \ , \qquad \Rightarrow \qquad \text{cov}(\beta u_t + \epsilon_t) \approx$$

The β controls banding. The size of δ determines tail dependence in the series. If δ is large then u_t^{δ} is near zero except for u_t near 1. Thus the effect of the noise ϵ_t is large except if u_t is near 1. In other words the two components in s_t are not related unless u_t is near 1. This situation is reversed if $\delta < 1$ and in particular if δ is near zero. In this case u_t near 1

Figure 1 illustrates

In the two dimensional case. In this case there is no loss in generality in assuming

$$s_{1t} = u_t \ , \qquad s_{2t} = \beta u_t^{\gamma} + \epsilon_t \ .$$

As $\beta \to \infty$ swamp any effect due to the normal error ϵ_t. Note that $\check{s}_{1t} = [u_t]$. When $\beta = 1$

In the two dimensional case

Figure xx illustrates things for the bivariate case. In this situation we take

Thus with Pareto marginals the model becomes

$$y_t = b\left\{(1-\check{s}_t)^{-1/a} - 1\right\} \ , \quad s_t = Z\alpha_t + \psi\cdot\epsilon_t \ , \qquad \epsilon_t \sim N(0,I) \ ,$$

where a and b are vectors of parameters and the first equation is interpreted in an obvious componentwise sense.

7.2 Other candidate marginal distributions

Since F_i^- maps numbers between 0 and 1 to real numbers it is convenient to think of them as

$$1'a + \sum_i b_i \ln\frac{u}{1-u} = 1'a + \ln\prod_i\left(\frac{\check{s}_{it}}{1-\check{s}_{it}}\right)^{b_i}$$

These have means a_i and variances $b_i^2\pi^2/3$.

8 Estimating loss distributions

This leads to claim distribution.

$$c_t = \sum_{i=1}^{p}\frac{b_i}{(1-\check{s}_{it})^{1/a_i}} - 1'b \ . \tag{8}$$

Simulating this loss is a three step process:

1. Simulating N realizations of s_t.
2. Computing the percentile ranks $\check{s}_t$.
3. Combining the $\check{s}_t$ as in (8).

This formula is of independent interest. If s_{it} is near 1, that is we are in the upper tail, then losses are likely to be huge.

Note that if u is uniform then

$$\frac{b}{(1-u)^{1/a}}$$

is Pareto In practice, for given parameter values, the losses ℓ_t and the loss distribution can be simulated. This yields relevant percentiles.

9 Time series structure

Up to now it is assumed the α_t are independent random variables. An alternative is where α_t has a time dependent structure. A simple example is a geometric random walk

$$\alpha_t = \mathrm{e}^{h_t/2} \ , \qquad h_{t+1} = h_t + \eta_t \ .$$

$$\alpha_{t+1} = (\eta_{t+1})^2$$

10 Connection to previous literature

This setup (4) is reminiscent of the usual factor model [Lawley and MaxwellLawley and Maxwell19 often employed to study the joint behavior of variables. Similar to the ordinary factor analysis Z spells out the way in which the heavy tailed factors manifest themselves into the components of s_t and hence their percentiles $\check{s}_t$.

The formulation (4) displays the obvious connection to state–space models [HarveyHarvey1989]. In terms of this latter model, α_t is called the state of the system which manifest in the signal s_t. The signal serves to determine the percentiles $\check{y}_t$ of y_t. The model is completed by specifying the marginal distributions of y_t.

11 Goodness of fit

Suppose n observations on a vector random variable y denoted $y_1, y_2, \ldots, y_n$. The aim is to estimate the parameters in Z and Ψ. Suppose these two matrices depend on unknown parameters θ and ρ, respectively.

The first step in the estimation is to convert the y_t to percentiles. This is done either relative to a fitted marginal distributions $\hat{F}$ in which case $\hat{F}(y_t)$ is the vector of estimated percentiles, or by computing the percentile ranks $\check{y}_t \equiv n^{-1}\varrho(y_t)$. In the former case the estimated percentiles need not be exactly uniformly distributed although this does hold in the latter case.

Suppose we use the $\check{y}_t$. Then according to the model $\check{y}_t = G(s_t)$ and

You can plot the time series of percentile ranks. Actually what you do is simulate a large number of Given a vector of actual or estimated marginal distributions F the percentile ranks are

$$\check{y}_t = F(y_t) \approx \frac{1}{n}\sum_{r=1}^{n}(y_r < y_t) \ , \qquad t = 1, \ldots, n \ .$$

The right hand side are the estimated percentile ranks in the situation if there is no estimate of F, and $(y_r < y_t)$ is the vector of 1's and 0's indicating whether the corresponding component of y_r is than y_t or otherwise.

The model states $\check{y}_t$ has the same distribution as $\check{s}_t$ where s_t is generated as in (4). For given θ and ρ, the percentiles $\check{s}_t$ are estimated by simulating a large number N of s_t from (4) and computing

$$\check{s}_t \approx \frac{1}{N}\sum_{r=1}^{N}(s_r < s_t) \ , \qquad t = 1, \ldots, N \ .$$

The distribution of the $\check{s}_t$ is compared to the distribution of the $\check{y}_t$. Good estimates of θ and ρ will lead to good correspondence between the outcomes $\check{y}_t$ and $\check{s}_t$.

One way to check whether the distribution of the $\check{s}_t$ is similar to that of the $\check{y}_t$ is to divide the unit hypercube into smaller volumes or cells indexed by c and count the number of $\check{y}_t$ in each cell. These are then compared to the expected number e_c calculated for given θ and ρ by simulating a large number of s_t and calculating the corresponding $\check{s}_t$. An appropriate choice for θ and ρ is indicated by a low value of chi–squared goodness of fit statistic

$$X^2 \equiv \sum_c \frac{(o_c - e_c)^2}{e_c} \ , \tag{9}$$

where the c sum extends over all cells. In practice X^2 in (9) is numerically minimized over a grid of values of θ and ρ.

12 Estimating total loss quantiles

Under the model (4) it is easy to simulate the distribution of total loss given the parameters λ and ρ. For each of the N simulated values of the state vector $s_t = Z\alpha_t + \Psi\epsilon_t$ the corresponding estimated loss is

$$\ell_t = 1'(\hat{F}^- \circ \hat{G})(s_t) = 1'\hat{F}(\check{s}_t) \ ,$$

where 1 is a vector of ones and $\hat{F}$ and $\hat{G}$ are the vectors of estimated marginal distributions. In particular $\hat{G}(s_t) = \check{s}_t$ where the latter is determined with reference to the N simulated s_t.

Ranking the $\hat{\ell}_t$, $t = 1, \ldots, N$ yields the $1/N$ quantiles of the loss distribution. These quantiles are conditional on the estimated copula parameters in λ and ρ and the assumed or estimated marginals in F. However under a suitably large sampling effort N, the quantiles will be independent of the approximation involved in G.

$$\ell_t = 1'F^-(\check{s}_t) \ , \qquad s_t = Z\alpha_t + \Psi\epsilon_t \ ,$$

where F is the vector of marginal distributions and F^- the vector of componentwise inverses. The marginal distributions F is either specified analytically, or determined empirically.

- This case requires proper spanning of the actual losses. Suppose y_t are the vectors of observed losses and $\check{y}_t$ the corresponding vector of percentile ranks. Then the total loss distribution is that of the ℓ_r, $r = 1, \ldots, n$ where

  $$\ell_r = 1'y_r \ , \qquad y_r \check{=} s_t \ , \qquad s_t = Z\alpha_t + \Psi\epsilon_t \ ,$$

  where the middle expression indicates the components of y_r are chosen from the components of the available y_t so that $\check{y}_r$ is as close as possible to $\check{s}_t$. Thus in this case a large number of s_t are simulated and converted to percentile ranks. Each observed y_r is then associated with a simulated s_t with the same, or approximately the same, percentile rank.

- In the case where F_* is available (either the true marginal distribution or an estimate) then the distribution of losses is built up from

  $$\ell_t = 1'F_*^-(\check{s}_t) \ , \qquad s_t = Z\alpha_t + \Psi\epsilon_t \ ,$$

  Thus in this situation a large number of s_t are simulated from the model from which the percentile ranks $\check{s}_t$ are computed. These $\check{s}_t$ are then used to compute ℓ_t corresponding to each simulation t.

Suppose y_t represents a vector of losses at time t on each of p classes of business. Then $x_t \equiv 1'y_t$ is the total loss and of interest is the quantile of x_t corresponding to the percentile $0 < \pi < 1$. In applications typical values for π are 0.95 and 0.99.

To estimate the quantile of x_t.

Consider first the two dimensional case. Percentile π in two dimensions is carved out of the unit cube as follows. Suppose firstly $y_{2t} = 0$. Then y_{1t} is the quantile corresponding to $\check{y}_{1t} = \pi$.

13 Iso–loss curves

If $0 < u < 1$ then u is called the p–value of the risk. A p–value close to 1 indicates it is extremely unlikely to get a total loss that exceeds the given value.

Suppose F is vector or marginal distribution functions and consider for vector $0 < u < 1$:

$$\ell(u) \equiv 1' F^-(u) \ .$$

This is the loss when each line of business is at the percentile corresponding to the component of u. Taking the total differential yields

$$\mathrm{d}\ell(u) = 1' \mathrm{d}F^-(u) = -1' \frac{\mathrm{d}u}{f\{F^-(u)\}} \ ,$$

Where $\mathrm{d}u$ is the vector of differentials and all the other operations are componentwise.

Paths where $\mathrm{d}\ell(u) = 0$ is called an iso–loss path since it defines changes in u leading to the same loss. If there are only two lines of business then $\mathrm{d}\ell(u) = 0$ is equivalent to

$$\frac{\mathrm{d}u_1}{f_1\{F_1^-(u_1)\}} = \frac{-\mathrm{d}u_2}{f_2\{F_2^-(u_2)\}} \qquad \Rightarrow \qquad -\frac{\mathrm{d}u_2}{\mathrm{d}u_1} = \frac{f_2\{F_2^-(u_2)\}}{f_1\{F_1^-(u_1)\}} \ .$$

This is the slope of an iso–loss curve and is the rate at which the percentile of one distribution must be changed for the percentile in another distribution while maintaining the same level of total loss. It connects all points in percentile space leading to the same loss and hence to which a company executive and hence regulator would be indifferent to. Maybe we can call this the "loss indifference curve."

Matters simplify when the loss distributions are Pareto. The univariate Pareto is, for parameters $\theta > 0$ and $\lambda > 0$,

$$F(y) = 1 - \left(\frac{\theta}{y+\theta}\right)^{1/\lambda} \ , \qquad F^-(u) = \theta\{(1-u)^{-\lambda} - 1\} \ , \qquad y > 0 \ .$$

This implies

$$f(y) = \frac{1 - F(y)}{\lambda(y+\theta)} \ , \qquad f\{F^-(u)\} = \frac{1-u}{\lambda\theta(1-u)^{-\lambda}} = \frac{u^{1+\lambda}}{\lambda\theta} \ .$$

Thus the total loss at each combination of percentiles is

$$F_1^-(u_1) + F_2^-(u_2) = \theta_1\{(1-u_1)^{-\lambda_1} - 1\} + \theta_2\{(1-u_2)^{-\lambda_2} - 1\} \ .$$

This is trivial to simulate given the copula $C(u)$. The above expression can be replaced by

$$\ell(u) = \sum_i \theta_i (u_i^{-\lambda_i} - 1) \ , \qquad \Rightarrow \qquad \mathrm{E}[f\{\ell(u)\}] = \int f \left\{ \sum_i \theta_i (u_i^{-\lambda_i} - 1) \right\} \mathrm{d}C(u)$$

Hence in this case, and supposing there are two components:

$$-\frac{\mathrm{d}u_2}{\mathrm{d}u_1} = \frac{\lambda_1\theta_1}{\lambda_2\theta_2}\left(\frac{u_2^{1+\lambda_2}}{u_1^{1+\lambda_1}}\right) \ .$$

Thus if the change in u_2 corresponding to a small change in u_1 is as given in the right hand side then the total loss is constant. What is the value of knowing this? Note if the shapes $\lambda_1 = \lambda_2 = \lambda$ then

$$-\frac{\mathrm{d}u_2}{\mathrm{d}u_1} = \frac{\theta_1}{\theta_2}\left(\frac{u_2}{u_1}\right)^{1+\lambda} = \mathrm{e}^{\ln\theta_1 - \ln\theta_2 + (1+\lambda)(\ln u_2 - \ln u_1)} \ .$$

Analogous to the economics literature this is called the marginal rate of percentile or risk substitution. From the expression on the right, on a log scale, the contours of constant total loss

To work with this suppose we take $u_1 = 0.95$ and $u_2 = 0.95$. The quantiles corresponding to these percentile are say $y_1(u_1)$ and $y_2(u_2)$ and total loss $y_1(u_1) + y_2(u_2)$. A small change $\mathrm{d}u_1 = 0.01$ to say $u_1 = 0.96$ leads to a larger y_1. Hence required is a small decrease in u_2 to arrive at the same total loss.

Should we say something about reinsurance arrangements?

The elasticity of percentile substitution is defined as

$$\sigma(u_1, u_2) \equiv \frac{\mathrm{d}\ln(u_2/u_1)}{\mathrm{d}\ln(-\mathrm{d}u_2/\mathrm{d}u_1)}$$

which is the percentage change in the relative size of the percentiles given a percentage change in the marginal rate of substitution. It is a measure of the ease with which the varying risk factors can be substituted for others. If it is measured with respect to $\ell(u)$ it measures the ease with which the varying risk factors can be substituted for others insofar as total loss is concerned. If it is measured with respect to $C(u)$ it measures the ease with which the varying risk factors can be substituted for others insofar as loss coverage probability is concerned. It effectively measures the curvature of an iso–loss or iso–probability curve.

When $\sigma \to \infty$ we have perfect substitutibility. If $\sigma = 0$ there is no substitutibility. For L shaped iso–probability curve $\sigma = 0$. This corresponds to $C(u) = \min u$ copula.

It would be nice to be able have a functional form for the components of $F_*(u)$ so that RQS take a simple form. That is we estimate some nice functional form

Suppose we have

$$\ell(u) = \left(\sum_{i=1}^{p} a_i^{\frac{1}{\sigma}} u_i^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \ . \tag{10}$$

Then the p marginal distributions are parametrized by the p numbers a_j and the constant elasticity of substitution σ.

To see the meaning of the a_i note $s(0) = 0$ while $s(1) = \sum a_i^{1/(\sigma-1)}$ which is the maximum loss. Further if u is all zeros except in component i where it is u_i then (10) reduces to

$$y_i = F_i^-(u_i) = a_i^{1/(\sigma-1)} u_i \qquad \Rightarrow \qquad u_i = F_i(y_i) = a_i^{1/(1-\sigma)} y_i$$

Thus $F_i(y_i)$ is linear in the tail up to some maximum point $y_i = a_i^{1/(\sigma-1)}$. Note that the maximum of $\ln y_i$ is $(\ln a_i)/(\sigma - 1)$.

Iso–loss lines can be easily constructed from

Given u we can easily determine the total loss to which this corresponds $s = 1'F^-(u)$. By simulating from the copula of u we can determine the frequency distribution of s. In particular if Ψ is the distribution of s then $G(s) = G\{1'F_*^-(u)\}$ is uniform. In other words $G \circ 1' \circ F_*^-$ is a mapping from copula distributed u to uniform random variables.

The rate of quantile substitution depends only on the relative magnitudes of the densities. The rates are used to construct a iso–loss contours. Integrating the density of copula above the iso–loss curve corresponding to s yields the probability of exceeding the loss s.

14 Iso–probability curves.

Copulas define contours on the unit hypercube called iso–probability curves. An iso–probability curve is the set of u such that $C(u)$ is constant. In particular the set of u such that $C(u) = 1 - \pi$, consists of all percentiles u such that

$$1 - \pi = C(u) = P\{y < F^-(u)\} \qquad \Rightarrow \qquad \pi = 1 - P\{y < F^-(u)\} \ .$$

In other words π is the exceedence probability, that is the probability that at least one component of y will exceed the the quantile (value–at–risk) corresponding to the appropriate component of u.

Denoting F as the vector of marginal distributions and the joint F_*, then

$$C(u) = (F_* \circ F^-)(u) = F_*\{F^-(u)\} = P\{y \le F^-(u)\} \ .$$

Hence u in $C^-(\pi)$ implies $P\{y \le F^-(u)\} = \pi$, meaning that if you take the u quantiles of y (vars) then the joint probability of being less than these quantiles (vars) is π.

The slope of an iso–probabiliy curve is called the marginal rate of probabiliy substitution (mrps) since it shows the reduction in the percentile in one

loss distribution sufficient to offset the increase in the percentile of another loss distribution so as to have exactly the same total loss probability. Note the mrps makes no reference to actual losses – only to percentiles. Note the scale problem – if one loss is in terms of \$bn and the other in terms of say \$'s then the mrps does not take cognisance of the relative scales – only coverage probability matters.

If losses are independent then $C(u) = \prod_j u_j$. Hence in the two dimensional case with independence the contours are $u_1 u_2 = \pi$ implying the slope is $-\pi/u_1^2$. In the completely dependent case $C(u) = \min u$ hence this corresponds to "no substitutability" and L shaped iso–probability curves.

15 Iso–loss curves

Consider the total loss $\ell(u) = 1' F^-(u)$ where u is a vector of percentiles. An iso loss contour is a set of u such that $\ell(u)$ is constant say c, and denoted $\ell^-(c)$.

The slope of an iso–loss contour is the marginal rate of percentile substitution indicates the rate at which the percentiles in one distribution must be changed when changing the percentiles in another distribution so as to keep the total loss constant. So relative scale is important.

Taking the total derivative

$$\mathrm{d}\ell(u) =$$

To illustrate, suppose we have two Pareto marginal distributions with the same shape:

$$F^-(u) = \theta(u^{-\lambda} - 1)$$

Generally both marginal rates of substitution are negative. However the mrltps is generally increasing while mrps is decreasing. This is explained as follows:

16 Optimal reserving

The optimal percentiles for reserving given c is u^* where:

- $\ell(u^*) = v$ is fixed. That is the value at risk is given by v.
- The marginal rate of percentile substitution with respect to loss is equal to the marginal rate of percentile substitution with respect to coverage probability

The value of $C(u^*)$ is the coverage probability and equal to the var percentile. The value $C(u^*)$ is the highest coverage probability given the total value at risk v. The common slope λ of the marginal rates of percentile substitution is the price of risk and indicates the marginal change in v by relaxing coverage probability.

Alternatively a point u^* can be viewed as the highest var $\ell(u)$ that can be obtained given the constraint $C(u) = c$. It is clear that $F^-(u^*)$ is the cost of capital allocation. Further $\ell(u^*) - \ell(\pi 1)$ is the diversification benefit where π is such that $C(\pi 1) = c$.

Alternatively

such that the two marginal rates of percentile substitution are the same. To see this, suppose equality does not hold at u. Travel along the iso–loss curve passing through u in an appropriate direction so as to move to the highest possible iso–probability curve, attained at point u^*, say. Then

$$\ell(u) = \ell(u^*) \ , \qquad C(u) \leq C(u^*) \ .$$

Thus the total loss has not changed but the probability of sufficient cover is higher. In other words setting aside $F^-(u^*)$ involves the same total reserve but there is more likely to be sufficient cover for each line.

Another way to look at things is to compute $\ell^-(c)$ as the set of u such that $\ell(u) = c$. Hence $\ell^-(u)$ defines an iso–quant where the quant is a monetary quantity of loss. Similarly let $C^-(\pi)$ as the set of u such that $C(u) = \pi$. Hence $C^-(\pi)$ defines an iso–percentile – that is a set of u with the same coverage probability.

Consider $\ell^-(c) \cap C^-(\pi)$, the intersection of these two sets. This set is either empty, contains one element, or contains two elements. Suppose U^* is the set of all u which is a one element intersection. This set is the optimal allocation problem in the following sense. Suppose $u^* \ \epsilon \ U^*$. Then $\ell(u^*)$ is the maximum loss corresponding to a given π. Alternatively $C(u^*)$ is the maximum coverage probability corresponding to a given loss c.

Hence plot $\ell(u^*)$ as a function of π or alternatively plot $C(u^*)$ as a function of c.

it contains one element then $u^*(c, \pi)$ be an element in this set. Want to take all the points such that the marginal rates of percentile substitution are the same.

17 Diversification benefit

The naive approach is to take reserve $\ell(\pi 1)$ where π is a scalar percentile. This states that reserves $F^-(\pi 1)$ are held with respect to each line of busi-

ness. This is represented as the diagonal line in the graphs. The coverage probability is $C(\pi 1)$ while the total reserve is $\ell(\pi 1)$.

This reserving strategy is optimal only in the case where the lines of business are independent in which case the sufficiency probability is $C(\pi 1) = \pi^p$ where p is the number of lines.

If lines are dependent then we can improve

What is the diversification benefit here? If we hold $F^-(u)$ reserves, then the total loss reserved is $\ell(u)$. The probability of simultaneously holding enough reserve for each line is $C(u)$. However by moving to u^* yields the same total loss reserve of $\ell(u^*) = \ell(u)$. However the probability of not exceeding any individual loss reserve is $C(u^*) > C(u)$. That is diversification has bought us the extra margin of security $C(u^*) - C(u)$.

Alternatively holding $F^-(u)$ where $u = \pi 1$ where π is scalar indicating quantile pi is held with respect to each line of business. This corresponds to taking a position on the diagonal. Now take a journey along the iso probability curve passing through diagonal point $u = \pi 1$ till attaining the lowest loss iso–quant at say u^*. For this u^*, $\ell(u^*) < \ell(\pi 1)$ but $C(u^*) = C(u)$. Hence the total loss is less but sufficient cover probability is the same. That is diversification saving is $\ell(u^*) - \ell(\pi 1)$ which is the difference in the iso–loss line. This is the diversification benefit measured in terms of monetary saving, keeping the loss probability the same.

Note u^* solves the capital allocation problem. That is $F^-(u^*)$ is the vector of capital amounts. Further u^* is the percentile allocation.

Should talk about the "loss coverage" at u. This is computed by computing firstly $\pi = C(u)$ and then finding the maximum $\ell(u)$ subject to $C(u) = \pi$.

Note under the independence copula

$$\ell(u) + \lambda C(u) = \sum_j F_j^-(u_j) + \lambda \prod_j u_j$$

In general terms the diversification benefit is

$$\max_u \ \ell(u) - \ell(\pi 1) \qquad \text{subject to} \qquad C(u) = C(\pi 1)$$

This is written as

$$\max_u \ \ell(u) - \ell(\pi 1) + \lambda\{C(u) - C(\pi 1)\} \ .$$

This leads to the first order conditions in terms of vectors of derivatives are.

$$\ell'(u) = \lambda C'(u)$$

The value of the Lagrangian λ_π is the price of risk indicating the extra saving to be obtained by relaxing the constraint.

Write ℓ_π as the solution. The solution traces out the risk profile on $[0,1]$ the diversification benefit on [0,1]. Of interest is the diversification benefit near 1. Alternatively can consider

$$x$$

Note iso–probability curves generally have increasing slope (ie are concave). Ditto for iso–loss curves. Hence equilibrium is likely to be unstable.

When the losses are perfectly dependent the iso–probability curves are L–shaped. Hence moving to the lowest loss curve means sticking to the diagonal. This suggests the diversification benefit is measured through the curvature of the iso–probability lines with maximum curvature attained under independence copula $\prod u_j$

Another way of saying this is to

Then by traveling along an iso–lower tail probability curve we retain the same lower tail probability but can move to lower iso–loss curve. In other words we can move to a situation with lower total loss and where the lower tail probability is the same.

at which mrqs is equal to mrps. This assumes we can reinsure the tails of each marginal distribution. If we are not at such a point then we can adjust u keeping the exceedence probability the same.

18 Coherent risk measures

Financial risks are finite random variables defined on some common probability space. If L is a financial risk then a number $\varrho(L)$ computed from L is called a risk measure. A coherent risk measure has the following properties:

1. For any scalar a, $\varrho(a+L) = a + \varrho(L)$.
2. For any scalar a, $\varrho(aL) = a\varrho(L)$
3. For any two risks, $\varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2)$
4. If $L_1 \leq L_2$ a.s. then $\varrho(L_1) \leq \varrho(L_2)$

Consider the π–quantile as a risk measure. Obviously 1, 2, and 4 are satisfied. Further consider the case where L_1 and L_2 are iid. Then from 3 and 2

$$\varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2) = 2\varrho(L_1) = \varrho(2L_1) \ .$$

Hence diversification across identical risks leads to a lower value for any coherent risk measure than loading up on the equivalent number of units of the same risk.

Now VaR is not coherent. This is because the π–quantile of a sum need not be less then the sum of π–quantiles. This can be illustrated with Bernoulli–bonds.

19 Expected shortfall

If F is a marginal distribution and $0 < u < 1$, then the expected shortfall at u is the expected loss given the loss exceeds the u–quantile:

$$s(u) \equiv \mathrm{E}\{L|L > F^-(u)\} = \frac{1}{1-u}\int_u^1 F^-(v)\mathrm{d}v \ .$$

The equality is proved using standard methods. It is proved in mcneil2005qrm that $s(u)$ is a coherent measure of risk. As u increases, so does $s(u)$.

The expected excess is defined as

$$s(u) - F^-(u) = \mathrm{E}\{L - F^-(u)|L > F^-(u)\}$$

As u increases, the net expected loss can both increase or decrease depending on tail behavior. With fat tails expected excess increases with u. Minimizing the sum of expected excesses will mean minimal not much u for those situations with fat tails.

Another criterion could be

$$(1-u)(s - F^-)(u)$$

This is the expected excess. As u increases the probability $1-u$ of an excess decreases even though the expected excess given there is excess may increase. In this case the aim is to

$$(1-u)'(s - F^-)(u) + \lambda\{\pi - C_*(u)\}$$

The above definition of $s(u)$ can be extended. As above all operations on vectors are interpreted in a componentwise sense. Note each component in $s(u)$ is increasing in the corresponding component in u.

But is $1's(u)$ a coherent measure of risk? To consider this we need to prove that a sum of coherent risk measures is itself coherent.

With the uniform distribution $s(u) = (1+u)/2 \to 1$ as $u \to 1$. Thus the problem of minimizing $C(u)$ subject to

The expression $s(u) - F^-(u)$ is the net expected shortfall. For the uniform this equals $(1+u)/2 - u = (1-u)/2$. Thus the net expected shortfall decreases to 0 as u increases to 1.

Why not minimize

$$1'(s - \delta F^-)(u) = 1'(s - F^-)(u) + (1 - \delta)1'F^-(u)$$

This is the net expected shortfall plus a reserving cost $(1 - \delta)1'F^-(u)$. Typically $1 - \delta$ reflects the opportunity cost of holding the reserve indicating a typical value for $\delta \approx 0.90$.

For the uniform

$$(s - \delta F^-)(u) = \frac{1 + u}{2} - \delta u = \frac{1 + (1 - 2\delta)u}{2}$$

If $u = 0$ net expected loss is $1/2$. As $u \to 1$ the net expected loss converges to $1 - \delta$.

Hence the minimization is

$$1'(s - \delta F^-)(u) + \lambda\{\pi - C_*(u)\} \ .$$

The first term captures the

Alternatively the maximization is

$$C_*(u) + \lambda\{c - 1'(s - \delta F^-)(u)\}$$

In this case we can maximize the probability of sufficiency while simultaneously reducing the expected shortfall to zero. So this is not realistic. Here expected shortfall is defined as $s(u) - F^-(u)$. Minimizing just $s(u)$ is not realistic either since

- Suppose $u^*(r)$ is the solution to $\max_u C_*(u)$ subject to $1'F^-(u) = r$. As the notation indicates the solution depends or r. This is the VaR criterion.

- If $s(u)$ is a componentwise function of u, consider

$$\max_u C_*(u) \qquad \text{subject to} \qquad 1's(u) = c \ . \tag{11}$$

  For example $s(u)$ can be expected shortfall as above in which case we are maximizing the probability of sufficiency subject to a constraint on the total expected shortfall. This is equivalent to

$$\max_y F_*(y) \qquad \text{subject to} \qquad 1'(s \circ F)(y) = c \ . \tag{12}$$

  If u^* is the maximizer to (11) then $y^* = F^-(u^*)$ is the solution to (12). The reserve is $1'y^* \equiv 1'F^-(u^*)$. Both problems are equivalent to

$$\min_u 1's(u) \qquad \text{subject to} \qquad C_*(u) = \pi \ . \tag{13}$$

  Particular instances of s are:

- $s = F^-$ implying the minimization of total VaR with respect to u subject to a constraint on probability of sufficiency.
- s as the expected shortfall given above, implying the minimization of total expected shortfall, again subject to a constraint on the probability of sufficiency.
- a vector of scaled identity functions, $s(u) = w \cdot u$ implying $1's(u) = w'u$. The vector w, say summing to 1 could indicate the relative importance of each risk. Thus in this case we are minimizing a convex linear combination of individual sufficiency probabilities subject to a constraint on the overall probability of sufficiency.

In all cases the reserve is $1'F^-(u)$. This suggests we should minimize the Lagrangian

$$\alpha 1's(u) + (1-\alpha)1'F^-(u) + \lambda\{\pi - C(u)\} \ .$$

The first term is the measure of risk, the second is the cost of the reserve position, while the third captures the constraint on the probability of sufficiency. Note if $s = F^-$ then this collapses to the constrained minimization of VaR. This also corresponds to $\alpha = 0$. If $\alpha = 1$ then total expected shortfall is minimized.

- The total differential of $1's(u)$ is, in the case of $s = F^-$,

$$1'\mathrm{d}s(u) = -1'\frac{\mathrm{d}u}{f\{F^-(u)\}} \ ,$$

where $\mathrm{d}u$ is the vector of differentials and all the other operations are componentwise. Paths where $\mathrm{d}1's(u) = 0$ are called iso–risk lines since these define changes in u leading to the same total risk as measured through s.

With two lines of business then $\mathrm{d}1's(u) = 0$ is equivalent to

$$\frac{\mathrm{d}u_1}{f_1\{F_1^-(u_1)\}} = \frac{-\mathrm{d}u_2}{f_2\{F_2^-(u_2)\}} \qquad \Rightarrow \qquad -\frac{\mathrm{d}u_2}{\mathrm{d}u_1} = \frac{f_2\{F_2^-(u_2)\}}{f_1\{F_1^-(u_1)\}} \ .$$

This is the slope of an iso–loss curve and is the rate at which the percentile of one distribution must be changed for the percentile in another distribution while maintaining the same level of total loss. It connects all points in percentile space leading to the same loss and hence to which a company executive and hence regulator would be indifferent to. Maybe we can call this the "loss indifference curve."

- The Lagrangian is
$$C(u) + \lambda\{r - 1's(u)\}$$
Differentiating with respect to u leads to the first order conditions:
$$\frac{\mathrm{d}C(u)}{\mathrm{d}u} = \lambda\frac{\mathrm{d}s(u)}{\mathrm{d}u} \ , \qquad r = 1's(u) \ ,$$
and λ is the marginal PoS of money. Note $\mathrm{d}s(u)/\mathrm{d}u$ is the vector of derivatives of the components of $s(u)$ with respect to the corresponding components in u.

- Alternatively consider the problem of minimizing the expected shortfall $1's(u)$ subject to a constraint on the probability of sufficiency. The Lagrangian is
$$1's(u) + \lambda\{\pi - C(u)\}$$

- In the two dimensional case the marginal rate of percentile substitution is defined as
$$\mathrm{d}C_*(u) = C'(u)\mathrm{d}u$$

- Suppose $s(u)$ is any componentwise vector of increasing functions of u. Set r such that $1's\{u^*(r)\} = w$ say. Then $s\{u^*(r)\}$ is the solution to maximizing $C(u)$ subject to $1's(u) = w$.

- Since
$$\mathrm{E}(L) = u\mathrm{E}\{L|L < F^-(u)\} + (1-u)\mathrm{E}\{L|L > F^-(u)\} \ .$$
This implies
$$(1-u)\mathrm{E}\{L|L > F^-(u)\} = \mathrm{E}(L) - u\mathrm{E}\{L|L < F^-(u)\} \ ,$$
and so
$$\int_u^1 F^-(v)\mathrm{d}v = \mathrm{E}(L) -$$

- Actually I think this definition should be changed to
$$s(u) - F^-(u) = \frac{1}{1-u}\left\{\int_u^1 F^-(v)\mathrm{d}v - (1-u)F^-(u)\right\} \ .$$

- Gaussian expected shortfall for a risk is
$$s(u) = \mu + \sigma\frac{\phi\left\{\Phi^-(u)\right\}}{1-u}$$
Hence for

- Pareto. In this case $F^-(u) = \theta(u^{-\lambda} - 1)/\lambda$ and hence

$$\frac{\theta}{1-u}\int_u^1 \frac{v^{-\lambda}-1}{\lambda}\mathrm{d}v = \frac{-\theta}{\lambda}+$$

  which is linear in θ. Hence the TES constraint is linear in the scale parameter.

- Note $1-C(u^*)$ is the probability of insufficiency. Hence $\{1-C(u^*)\}\{1's(u^*)\}$ is an estimate of the expected insufficiency.

- The Pareto distribution depends on parameter values $\gamma > 1$ and $\theta > 0$ and has the form, for severity $x > 0$,

$$\mathrm{P}(L > x) = \left(\frac{\theta}{x+\theta}\right)^\gamma , \qquad \mathrm{E}(H^*) = \frac{\theta}{\gamma-1}\left(\frac{\theta}{d+\theta}\right)^{\gamma-1} ,$$

  with $d = 0$ yielding $H^* = H$. If d is such that $\mathrm{P}(H > d) = \delta$, then proportion $1-\delta$ of head injuries are reduced to zero and

$$\mathrm{P}(H^* > 0) = \left(\frac{\theta}{d+\theta}\right)^\gamma = \delta \qquad \Rightarrow \qquad h^* = \delta^{1-1/\gamma}h ,$$

E(

Now suppose we have two risks and u_1 and u_2 are percentiles. Consider the iso–shortfall curve $c = s_1(u_1) + s_2(u_2) = 1's(u)$. This is the set of all percentiles with the same total expected shortfall.

If expected shortfall (ES) is the risk measure then an equilibrium risk management position (u_1^*, u_2^*) is where for given PoS (probability of sufficiency) we have the lowest ES. Alternatively for given ES we have the highest PoS. This means the mrps (PoS) equals the mrps (ES). The capital allocation is then $s(u^*)$, the vector of expected shortfalls at the optimizing u^*. The price of capital is the common mrps.

20 Distortion operators

First note for a positive random variable L and $0 < u < 1$,

$$(1-u)\mathrm{E}(L|L > y_u) = \int_{y_u}^\infty yf(y)\mathrm{d}y = y_u + \int_{y_u}^\infty \{1-F(y)\}\mathrm{d}y = \int_u^1 F^-(v)\mathrm{d}v ,$$

where y_u is the u–quantile of y. The first equality is follows from integration by parts. Defining $v = F(y)$ then $\mathrm{d}v = f(y)\mathrm{d}y$ and $y = F^-(v)$ and substituting into the first expression yields the third expression.

Write $S(y) \equiv 1 - F(y)$ and consider

$$\int_0^\infty (\phi \circ S)(y) \mathrm{d}y = \int_0^\infty y \phi'\{S(y)\} f(y) \mathrm{d}y \ .$$

where the right hand side is established using integration by parts and assumes the derivatives ϕ' and $-f$ of respectively ϕ and S exist. Then

- If $\phi(x) = 1$ then the integral diverges.
- If $\phi(x) = x$ then the integral equals $\mathrm{E}(L)$.
- If $\phi(x) = 0$ if $x \leq 1 - u$ and 1 otherwise then the integral is q_u, the u–quantile of L since in this case the left hand side expression is

  $$\int_0^{q_u} 1 \mathrm{d}y = q_u \ .$$

  Further, in this case, ϕ' does not exist and the right hand side expression above is ill defined.
- The conditional tail expectation results if $\phi(x) = x/(1-u)$ if $x \leq 1-u$ and 1 otherwise since in this case the above expression becomes

  $$\int_0^{q_u} 1 \mathrm{d}y + \frac{1}{1-u} \int_{q_u}^\infty S(y) \mathrm{d}y = q_u + \mathrm{E}(L | L > q_u) \ .$$

  xx shows tail conditional expectations can be written as
  Now "distort" things:

$$\frac{1}{1-u} \int_{y_u}^\infty (\phi \circ F)(y) \mathrm{d}y = \frac{1}{1-u} \int_u^\infty (F^- \circ \phi^-)(v) \mathrm{d}v \ .$$

Note ϕ is a function from [0,1] to [0,1]. Now $(\phi \circ F)^- = F^- \circ \phi^-$ since $y = (\phi \circ F)^-(v)$ implies $\phi\{F(y)\} = v$ implying $F(y) = \phi^-(v)$ and in turn $y = F^-\{\phi^-(v)\}$, as required. Hence insofar as tail metrics are concerned, distorting F with ϕ is equivalent to first distorting the uniform by ϕ^- and then applying F^-.

Distortion operators serve to emphasize or deemphasize the tail of the distribution. Thus the problem becomes:

$$\frac{1}{1-u} \int_v^1 1'(F^- \circ \phi^-)(v) \mathrm{d}v + \lambda\{\pi - C(u)\}$$

Thus the reserve requirement is "shaped" by the distortion operator ϕ.

Can we find a distortion operator that handles the summation nicely. That is suppose the sum of the F^- is turned into something really nice. The nicest thing would be where you sum across a linear combination. ie It would be nice if the sum of $\sum_j (F_j^- \circ \phi_j^-)(u_j) = \sum_j c_j u_j$

21 Risk management

Suppose that a risk manager maximizes PoS subject to a constraint on the total expected shortfall (TES). The Lagrangian expression for this problem is

$$C(u) - \lambda\{1's(u) - \tau\} \ .$$

Note that in the Pareto case the TES constraint is linear in the scaling constant θ. Maximization requires the solution to

$$\mathrm{d}C(u) = \lambda \mathrm{d}s(u) \ , \qquad 1's(u) = s_0 \ .$$

Note $\mathrm{d}C(u)$ is the vector of partial derivatives of $C(u)$ with respect to each of the u. Further $\mathrm{d}s(u)$ is the vector of derivatives of each component of $s(u)$ with respect to the corresponding component in u. Here λ is the marginal PoS given an extra unit of TES. Obviously the above is the same as minimizing TES subject to a constraint on PoS.

The above equations hold at any optimum u^*. As s_0 and the parameters θ in $s(u)$ vary, the u^* and λ^* vary but the equations still hold. Thus u^* and λ^* are functions of θ and λ:

$$u = u(\theta, \lambda) \ , \qquad \lambda = \lambda(\theta, s_0) \ .$$

When these percentile functions are substituted into the above equations the quantities on the left become identically equal to 0 in θ and λ. Therefore the partial derivatives of these quantities with respect to θ_1 is zero:

$$\begin{pmatrix} \frac{\mathrm{d}^2 C}{\mathrm{d}u \mathrm{d}u'} & -\frac{\mathrm{d}s}{\mathrm{d}u} \\ -\frac{\mathrm{d}s}{\mathrm{d}u}' & 0 \end{pmatrix} \begin{pmatrix} \frac{\mathrm{d}u}{\mathrm{d}\theta_1} \\ \frac{\mathrm{d}\lambda}{\mathrm{d}\theta_1} \end{pmatrix} =$$

Solving yields

$$\frac{\mathrm{d}u_i}{\mathrm{d}\theta_1} = \lambda r_{i1} + r_{i,m+1} \frac{\mathrm{d}s_1(u_1)}{\mathrm{d}\theta_j}$$

where the r's are the elements of the inverse of the matrix on the left–hand side. Further $\mathrm{d}u_i/\mathrm{d}s_0 = -r_{i,m+1}$. Hence this last derivative is called the constraint effect of the parameter change and the first term above the "subsititution" effect.

In other words when a parameter is changed in a marginal distribution the optimum u^* changes partly by substituting away from or into the relevant risk and partly on account on the constraint effect.

The above analysis could be useful if one wanted to see how sensitive a risk position is to changes in the assumed marginal distributions. Here are some notes to keep in mind:

- The loss distribution
-
-

22 Copulas from total loss

Suppose $\ell_*(u)$ denotes the loss if at percentiles u on the individual portfolios. Normalize so that the maximum loss is $\ell(1) = 1$. Note $\ell_* = 1' \circ F^-$ and define for scalar $0 \leq \pi \leq 1$,

$$\ell_i(\pi_i) \equiv \ell_*(1, \ldots, \pi_i, \ldots, 1) = 1 - F_i^-(1) + F_i^-(\pi_i) \ .$$

where $0 \leq \ell_i(\pi_i) \leq 1$. Hence $\tau_i = \ell_i(\pi_i)$ is the total loss if all lines are at max loss except line i which is at percentile π_i. Further $\pi_i = \ell_i^-(\tau_i)$ is the percentile loss in line i if the total loss is τ_i and all other lines are at maximum loss.

Now consider the "copula" generated from ℓ_*:

$$\ell_* \circ \ell^- = 1' \circ F^- \circ \ell^- = 1' \circ (\ell \circ F)^- \ .$$

Now $\ell \circ F$ computes from a vector of losses y the vector max losses where each component assumes, all other losses are at their maximum. Note the max loss vector evaluated at the vector 0 is the max loss excluding line i. Thus $(\ell \circ F)^-$ computes from a max loss vector, the vector of individual losses y. In words, given a vector of total losses where each component corresponds to the maximum loss on all other lines and the current line having loss y_i.

Suppose $\ell_*(u)$ denotes the loss if at percentiles u on the individual portfolios. Normalize so that the maximum loss is $\ell(1) = 1$. Note $\ell_* = 1' \circ F^-$ and define for scalar $0 \leq \pi \leq 1$,

$$\ell_i(\pi_i) \equiv \ell_*(0, \ldots, \pi_i, \ldots, 0) = F_i^-(\pi_i) \ .$$

where $0 \leq \ell_i(\pi_i) \leq 1$.

Now consider the "copula" generated from ℓ_*:

$$\ell_* \circ \ell^- = 1' \circ F^- \circ \ell^- = 1' \circ (\ell \circ F)^- \ .$$

Now $\ell \circ F$ computes from a vector of losses y the vector max losses where each component assumes, all other losses are at their maximum. Note the max loss vector evaluated at the vector 0 is the max loss excluding line i. Thus

$(\ell \circ F)^-$ computes from a max loss vector, the vector of individual losses y. In words, given a vector of total losses where each component corresponds to the maximum loss on all other lines and the current line having loss y_i.

Is there a nice formula for $(\ell \circ F)^-$. The operator maps a max loss vector to a loss vector.

What is the joint distribution of max loss vectors? That is suppose I generate loss y_i.

23 Further stuff

A sampling scheme is as follows.

$$s = 1'\hat{F}_*^-(u) \ , \qquad u \sim C$$

Alternatively to find out what quantile a given s corresponds to, first construct the iso loss line corresponding to s. Repeatedly draw u and see the proportion of those u for which $1'\hat{F}_*^-(u)$ exceed s.

An improvement is where we first select u_1 uniformly from $(0, 1)$. If $y_1 \equiv F_1^-(u_1) > s$ then stop or alternatively if $F_1(s) > u_1$ then stop. If not compute draw $u_2 \sim u_2|u_1$ and $F_2(s - y_1) > u_2$ then stop.

the positive number $s - F_1^-(u_1)$. This is the number $y_2 = F_2^-(u_2)$ has to exceed in order for the sum to exceed ... That is we want

$$F_2\{s - F_1^-(u_1)\}$$

Accept–Reject sampling can also be used. We draw

Consider $p_1u_1 + p_2u_2$ where p_1 and p_2 are prices. The prices may represent the cost of reducing each of the quantiles. To minimize s subject to

In the two dimensional case.

The aim is, to find λ and ρ consistent with the observed percentile ranks $y_j^\#$. This can be done as follows.

24 Turning the loss surface into a copula

Given the marginals F, then the total loss at u is $\ell(u) = 1'F^-(u)$. Normalize things so that $\ell(1) = 1$. Note if all components of u are set to zero except component i then we recover $F_i^-(u_i)$. If all components of u are set to 1 except component i then we get

$$1 - F_i^-(1) + F_i^-(u_i) = 1 - \{m_i - F_i^-(u_i)\}$$

where $m_i \equiv F_i^-(1)$, the maximum loss on line i.

Now consider

$$M(z) \equiv \mathrm{E}\left\{z^{\ell(u)}\right\} = \int z^{\ell(u)} \mathrm{d}C(u) \ .$$

Then $M(z)$ is the mgf of the total loss and

$$M(0) = 0 \ , \qquad M(1) = 1 \ , \qquad M'(1) =$$

and M behaves like a distribution function. Note that in this description $\ell(u)$ is strictly determined by the marginals while $C(u)$ strictly determines the dependence between losses.

24.1 Importance sampling

Say we want to estimate the 95% quantile of the total loss distribution. That is we want ℓ_q such that

$$P\{\ell(u) < \ell_q\} = q$$

The left hand side

Suppose things are scaled such that $1'F^-(1) = 1$. Then $H_* = 1' \circ F^-$ is a probability distribution on the unit cube. It represents the fraction of maximum loss when at u. The marginals associated with this probability distribution are

$$H_i(u_i) = (1' \circ F^-)(1, \ldots, u_i, \ldots, 1) = 1 - F_i^-(1) + F_i^-(u_i) = m_i + F_i^-(u_i) \ ,$$

where $m_i = 1 - F_i^-(1)$ is the maximum loss on all portfolios excluding i. Further

$$H_i^- = (m_i + F_i^-)^- = F_i - m_i \qquad \Rightarrow \qquad H_* \circ H^- = (1' \circ F^-) \circ (F - m) = \ ,$$

$$H_* \circ H^- = (1' \circ F^-) \circ (F - m) = 1' \circ (I - F^- \circ m) \ ,$$

where m is the vector of m_i.

Note $F^- \circ m$ has entry i equal to $F_i^-(m_i)$, the total loss on portfolio i when .... . Hence $H_* \circ H^-$ is a copula generated by the marginal distributions H. Note

Hence this copula just looks at how far each

Furthermore

$$H = (H_* \circ H^-) \circ H \ , \qquad F = \{M_*(1 - H) - M\}^-$$

25 The framework

Suppose L_1 and L_2 are two loss variables with marginal distributions F_1 and F_2 and respective inverses F_1^- and F_2^-. Suppose $C(u_1, u_2)$ is the copula associated with the joint distribution of L_1 and L_2.

- An iso–loss curve is the set of all u_1 and u_2 such that $F_1^-(u_1) + F_2^-(u_1)$ is constant.
  - Generally iso–loss curves are concave.
- An iso–pos curve is the set of all u_1 and u_2 such that $C(u_1, u_2)$ is constant
  -
  - Generally iso–pos curves are convex.
-

26 Junk

Thus we have $C_{\lambda,\rho}$ and our model is

$$y \sim C_{\lambda,\rho} \circ F_* \ , \qquad y \sim F_*^- \circ C_{\lambda,\rho}$$

Note the left hand side states how probabilities associated with y are calculated, namely

$$P(y \leq \tilde{y}) = (C_{\lambda,\rho} \circ F_*)(\tilde{y}) \equiv C_{\lambda,\rho}\{F_*(\tilde{y})\} \ ,$$

while the right hand side indicates how one generates y: first you generate the uniforms according the copula, then you use these uniforms to interpolate out of the marginals: namely

$$y = F_*^-(u) \qquad \text{where} \qquad u \sim C_{\lambda,\rho} \ .$$

Alternatively we can thus write the setup as

$$y = F_*^-(s^{\#}) \ , \qquad s = \Lambda f + D\epsilon \ .$$

27 Connection to time series models

A time series version of this is

$$\check{y}_t = \check{s}_t \ , \qquad s_t = Z\alpha_t + \Psi u_t \ , \qquad u_t \sim N(0, I) \ , \qquad \alpha_t \sim \chi^2_1 \ .$$

These generated s values are used to find synthesized $s(y_j)$ values corresponding to the y_j such that $s^{\#}(y_j) = y_j^{\#}$. We then measure

$$\min_{\psi} \sum_{j=1}^{n} \|s_j - s(y_j)\|^2$$

and construct the empirical distribution of the percentile ranks $s_j^{\#}$, $j = 1, \ldots, N$. Corresponding to each $y_j^{\#}$, pick out $s(y_j^{\#})$, which is the generated s with percentile ranks $s^{\#}$ closest to that of $y_j^{\#}$. Then the log–likelihood of (λ, ρ) based on $y_j^{\#}$, $j = 1, \ldots, n$ is

$$\ell(\lambda, \rho) \approx \sum_{j=1}^{n} \ln f\{s(y_j^{\#})|\lambda, \rho\} \ ,$$

where here f

28 Copula properties

Consider a multivariate probability distribution F. Define F_* as the vector of marginal distributions derived from F. In particular if $y \equiv (y_1, y_2, \ldots, y_p)$ then

$$F_*(y) \equiv \begin{pmatrix} F(y_1, \infty, \ldots, \infty) \\ F(\infty, y_2, \ldots, \infty) \\ \vdots \\ F(\infty, \ldots, \infty, y_p) \end{pmatrix} \equiv \begin{pmatrix} F_1(y_1) \\ F_2(y_2) \\ \vdots \\ F_p(y_p) \end{pmatrix} \ .$$

In other circumstances the vector of marginals F_* is specified directly.

A copula is a probability distribution C on the unit hypercube with uniform marginals. Thus $C_*(y) = y$ for $0 \leq y \leq 1$ implying $C_* = 1$.

The usefulness of copulas, arises as follows. Consider a multivariate distribution F and define the distribution generated with copula C and given marginals F_*:

$$F_C \equiv C \circ F_* \ , \tag{14}$$

Then F_C is a joint distribution with $(F_C)_* = F_*$ since F_C has marginals

$$(C \circ F_*)(\infty, \ldots, y_i, \ldots, \infty) = C\{1, \ldots, F_i(y_i), \ldots, 1\} = C_i\{F_i(y_i)\} = F_i(y_i) \ .$$

Conversely given a joint distribution F with marginals F_* and F_*^- as the componentwise inverse to F_* define the copula generated from F as

$$C_F \equiv F \circ F_*^- \ . \tag{15}$$

Then C_F is a copula since it has uniform marginals:

$$(C_F)_i(u_i) = (F \circ F_*^-)(1, \ldots, u_i, \ldots, 1) = F\{\infty, \ldots, F_i^-(u_i), \ldots, \infty\} = u_i \ .$$

Every F has a copula representation which is clear from

$$F_{C_F} = C_F \circ F_* = F \circ F_*^- \circ F_* = F \ .$$

Finally note

$$C_C = C \ , \qquad C_{F_C} = F_C \circ (F_C)_*^- = C \circ F_* \circ F_*^- = C \ .$$

29 Copula properties

- $C_*(y) = y$. If $y \sim F_C$ then $F_*(y) \sim C$. If $u \sim C_F$ then $F_*^-(u) \sim F$.
- If H_* is the componentwise vector of univariate monotonic functions then
$$C_{F \circ H_*^-} \equiv F \circ H_*^- \circ (F \circ H_*^-)_*^- = F \circ F_*^- \equiv C_F \ .$$
Hence copulas are invariant to monotonic transformations.
- The density associated with $C_F = F \circ F_*^-$ is
$$\frac{f \circ F_*^-}{\Pi(f_* \circ F_*^-)} \ ,$$
where f is the density associated with F and f_* is the vector of marginal densitities. If $f = \Pi f_*$ (ie independence) then $f \circ F_*^- = \Pi(f_* \circ F_*^-)$ and the density is identically 1.
- The Gaussian copula with correlation matrix R is the copula C_{Φ^R} where Φ^R is the Gaussian distribution with mean 0 and covariance matrix R. Since R is a correlation matrix $(\Phi^R)_* = \Phi_*$, a vector of standard normal distributions, and $C_{\Phi^R} = \Phi^R \circ \Phi_*^-$. The meta–Gaussian model is
$$F_*(y) \sim C_{\Phi^R} \qquad \Leftrightarrow \qquad F_*(y) = \Phi_*(\epsilon) \ , \qquad \epsilon \sim N(0, R) \ . \tag{16}$$
Matrix R is estimated using the average cross product matrix of the normits $(\Phi_*^- \circ F_*)(y)$.

- The meta–gamma model is

$$F_*(y) = \Gamma_\psi(1\eta + \Psi\epsilon) \ , \qquad (\eta, \epsilon')' \sim \Gamma(\mu, \nu) \ ,$$

  where ψ is a vector of parameters, $\Psi = \text{diag}(\psi)$ and Γ_ψ the vector of marginal distributions associated with $1\eta + \Psi\epsilon$.

- The Levy copula is L such that $L_*(u) = u$. Thus it is a copula on the unit hypercube suitable extended to $u > 1$.

- The independence and perfectly dependent copulas is defined as $\Pi \equiv \Pi\, 1_*$ and $\vee \equiv \min$ respectively, where 1_* indicates the vector of identity functions. These arise as follows:

$$C_{\Pi F_*} \equiv (\Pi F_*) \circ F_*^- = \Pi\ 1_* \ , \qquad C_\vee = \ .$$

  The perfectly dependent copula implies a single source of noise: $y = F_*^- 1u$ with $u \sim U$ where 1 is a vector of identity functions of one variable.

- Any copula is "symmetric" in the sense that if C is the copula of 2D vector u then it is also the copula of the u with the components interchanged. This follows since suppose the density associated with C is c. Then since $c(u_1) \equiv 1 \equiv c(u_2)$,

$$c(u_1|u_2) = \frac{c(u_1, u_2)}{c(u_2)} = c(u_1, u_2) = \frac{c(u_1, u_2)}{c(u_2)} = c(u_2|u_1)$$

- If C is a copula then $C \circ (1-)$ is the "survival" copula. Here $(1-)$ is defined such that $(1-)u = 1 - u$. To prove this result note that

30 Factor copula models

The simplest factor copula model is as follows. Given a vector of nonnegative weights $\psi \geq 0$, suppose G is the distribution of

$$1\eta + \Psi\epsilon \qquad \text{where} \qquad \eta \sim \chi_1^2 \ , \qquad \epsilon \sim N(0, I)$$

with η and ϵ independent and $\Psi = \text{diag}(\psi)$. Further suppose G_* is the vector of marginal distributions associated with G. Then $u \equiv G_*(1\eta + \Psi\epsilon)$ is a vector of uniform, albeit dependent, random variables. Hence the distribution of u is a copula.

In terms of this development we suppose y is generated by the marginals F_* and copula derived from G if

$$F_*(y) = G(\Lambda f + \Psi\epsilon) \qquad \Leftrightarrow \qquad y \sim F_*^- \circ C_G \ .$$

These steps make it clear how things are structured. First, the factor scores are generated. Then transformed to uniforms and finally transformed from the uniforms to the things in the tails.

31 Generated practical copula structures

To simulate the model for given F_* and given ψ, we first generate the many values of the scores $1f + \Psi\epsilon$. These scores are then transformed to percentile ranks and these percentile ranks are used to pull out the corresponding values from the marginal distribution.

Initially suppose $f = \chi_1^2$.

$G_*(1f+\Psi\epsilon)$ one first simulates $1\eta+\Psi\epsilon$ and then applies the G_*^ψ. However a simpler, more practical procedure is simply to simulate a large number of $1\eta+\Psi\epsilon$ and then compute empirical quantiles corresponding to each component. Figure **??** displays realizations in the bivariate case for four different values of ψ. In each case the first component of ψ is zero while the second component is the given positive value.

Each of these figures was generated as follows.

1. For $t = 1, \ldots, n = 10000$ generate $\eta_t \sim \chi_1^2$ and $\epsilon_t \sim N(0, I)$ of order 2 and construct $\gamma_t \equiv 1\eta_t + \text{diag}(\psi)\epsilon_t$, $t = 1, \ldots, n$.

2. Compute the quantiles associated with γ_{it} for each $i = 1, 2$.

32 Wang article

Suppose x is a vector of losses (random). Assume from x you derive (componentwise) a vector of risk capitals $r(x)$ where r works componentwise. There are two rules for calculating capital:

- Basel Accords: $\iota' r(x) = \sum_i r_i(x) = \sum_i r_i(x_i)$.
- US Insurance NAIC RBC: $\sqrt{\sum_i r_i^2(x_i)}$.
- Company models: Individual probability distributions for a company's risks are quantified. Then assuming a correlation matrix the aggregate

loss distribution is calculated yielding percentiles etc. The total economic capital is then distributed back to the business units according to

$$\pi_i = \frac{\iota' R u_i}{\iota' R \iota} \ , \qquad \iota = \sum_i u_i$$

where u_i is a vector of zeros except in position i where it is 1.

Next states that yield spreads exaggerate default rates. Goes on to posit the model for the price of risk h_t where

$$h_t = \hat{\mu}_t + \lambda_t \sigma_t \ , \qquad \hat{\mu}_t = \mu_t + \epsilon_t \ , \qquad \mathrm{cov}(\epsilon_t) = (\partial\sigma_t) C (\partial\sigma_t)$$

where λ_t is a scalar stochastic process, C is a correlation matrix and $\partial\sigma_t$ is the diagonal matrix formed from the standard deviation vector σ_t. Also μ_t is the true mean value of the loss: ie $\mu_t = \mathrm{E}(x_t)$. Assume $\mathrm{cov}(\epsilon_t, \lambda_t) = 0$ implying

$$\mathrm{cov}(h_t) = (\partial\sigma_t) C (\partial\sigma_t) + \mathrm{cov}(\lambda_t) \sigma_t \sigma_t'$$

32.1 Corner tail correlations

The proposed measure of tail correlation proceeds as follows. take the percentile ranks and map to the z–scores ie $z_j = (\Phi^{-1} \circ F_*)(x_j)$. Assuming x_j is bivariate, write $\pi(\alpha)$ as the proportion of the x_j falling in the lower corner. The lower tail correlation is the solution for the off diagonal entry in the correlation matrix R of

$$(\Phi^R \circ \Phi^{-1})(1\alpha) = \mathrm{E}[\Pi\{\hat{F}_*(x_j) < 1\alpha\}] \ .$$

where E denotes the empirical expectation (ie fraction). Why not try and find the optimal R.

33 Copula factor model

Suppose P computes the "uniforms" of a random variable. Then for any monotonic transform f, $\mathrm{P} \circ f = \mathrm{P}$ and hence $\mathrm{P} \circ \mathrm{P} = \mathrm{P}$. Further for a monotonic decreasing function $\mathrm{P} \circ f = 1 - \mathrm{P}$ implying $\mathrm{P} \circ (1 - \mathrm{P}) = 1 - \mathrm{P}$. Also $\mathrm{P} \circ - = 1 - \mathrm{P}$.

And suppose ψ a vector of weights. Then y is said to be generated by the copula factor model if

$$\mathrm{P}(y_t) = \mathrm{P}(\Lambda f_t + \Psi \epsilon_t) \ , \qquad f_t \sim \chi^2_\nu \text{ or } t_\nu \ , \qquad \epsilon_t \sim N(0, \sigma^2 I) \ ,$$

where Ψ is a diagonal matrix with diagonal ψ. Without loss of generality assume the first component of ψ is zero.

33.1 Different specifications

Here show the effect of different Λ and Ψ and perhaps replacing $\Lambda f_t + \Psi\epsilon_t$ by eg $\Lambda(f_t) + \Psi\epsilon_t$.

33.2 Rank correlation

Suppose $\mathrm{R}(y_t)$ is the rank correlation matrix of y_t: that is $\mathrm{R}(y_t) \equiv \mathrm{cor}\{\mathrm{P}(y_t)\}$

33.3 Estimating the copula factor model

Consider $\hat{\mathrm{P}}(y_t)$. If we know the cdf P_x of $x_t \equiv \Lambda f_t + \psi\epsilon_t$ then $\hat{x}_t = (P_x^- \circ \hat{P})(y_t)$. However we don't know P_x. However simulate x_s, $s = 1, \ldots, N$ and reorder the simulated components of x_s, $s = 1, \ldots n$ to yield $\tilde{x}_t$ so that $\hat{\mathrm{P}}(\tilde{x}_t) = \hat{\mathrm{P}}(y_t)$. In other words we order the simulated components of x_s according to the order specified in y_t.

What is the relationship of this to finding ψ such that the rank correlation is preserved. In this we simulate x_t from the model and adjust ψ till we get a rank correlation for the x_t equal to the observed rank correlation. Obviously this procedure cannot tell us anything about order of things.

So we use the procedure of the last paragraph to find ψ. Then for the given ψ we simulate x_s, $s = 1, \ldots, n$ and in turn $\tilde{x}_t$. We use the $\tilde{x}_s$ to infer about the f_t.

Given the estimate $\tilde{x}_t$ of $x_t = \Lambda f_t + \Psi\epsilon_t$ we can do a PC analysis on $\tilde{x}_t$.

Then $\hat{\mathrm{P}}(\tilde{x}_t) = \hat{\mathrm{P}}(y_t)$ and $\tilde{x}_t$ is our estimate.

and in particular $+/\hat{\mathrm{P}}(y_t)$. If we know Λf_t then take

Consider the the principal components of $\hat{\mathrm{P}}(y_t)$. These pick up the major directions of variation. The vector ψ is estimated by the following steps:

1. Calculate the principal components of $\hat{\mathrm{P}}(y_t)$ to yield say UDU'.

2. If u is the first column of U compute $u'\hat{\mathrm{P}}(y_t)\} \approx \mathrm{P}(f_t)$. This the vector of weights.

3. Minimize

Given ψ and hence Ψ we simulate $1f_t^s + \Psi\epsilon_t^s$ and reorder each scalar series in each simulation to yield x_t^s such that $\hat{\mathrm{P}}(x_t^s) = \hat{\mathrm{P}}(y_t)$. Note that x_t^s is y_t interpolated back into the EDF of $1f_t^s + \Psi\epsilon_t^s$.

33.4 The two variable case

To illustrate estimation consider first the case where y_t is two dimensional $\Lambda = 1$. Suppose r is the correlation between the ranks the two components in $\hat{\mathrm{P}}(y_t)$ – that is the rank correlation between the components in y_t. Then we want to estimate ψ subject to

$$\mathrm{cor}\{\hat{\mathrm{P}}(1f_t + \Psi\epsilon_t)\} = \mathrm{cor}\{\hat{\mathrm{P}}(y_t)\} \ .$$

Notice the correlation decreases with ψ. So we simulate $1f_t + \Psi\epsilon_t$ and compute $\mathrm{cor}\{\mathrm{P}(1f_t + \Psi\epsilon_t)\}$. Then

Put y as the $2 \times n$ vector of y_t, $t = 1, \ldots, n$. Write $x_t = \Lambda f_t + \Psi\epsilon_t$ and without loss of generality $\Psi = \mathrm{diag}(0, 1)$. For given $s = \sigma$ generate a simulated path of x_t denote x^s. Note x^s is $2 \times n$. Rank each row of x^s according to the ranks of y denoted $\tilde{x}^s$ and redefine s as the standard deviation of $\Delta\tilde{x}^s$ where the latter subtracts the first row from the second. Cycle through this procedure as in Gibbs sampling to get a distribution of s. This is the posterior distribution of σ.

Then calculate the units of each component in y_t and calculate its value in the empirical cdf of the corresponding component of x_t. To achieve this rank, for each i the x_{it}, $t = 1, \ldots, n$ according to the ranks of y_{it}, $t = 1, \ldots, n$. Suppose this yields $\tilde{x}_t$. This is equal to ϵ_{2t}. So take the standard deviation of so as to get an estimate of σ.

Thus the joint distribution is generated with the meta–gamma copula $C^\psi \equiv G^\psi \circ (G^\psi_*)^-$.

Maybe the model should be changed to something like this

$$F_*(y) = H_*[1\Phi^-\{G(\eta)\} + (\partial\psi)\epsilon] \ , \qquad \eta \sim \chi^2_1 \ , \qquad \epsilon \sim N(0, I) \ .$$

33.5 Estimation of copula parameters

Estimation and testing consists of the following steps:

1. Compute the empirical cdfs $\hat{F}_*(y)$ and the percentile ranks $u_t = \hat{F}_*(y_t)$.
2. Draw $\eta_i \sim \chi^2_\nu$ and $z_i \sim N(0, 1)$ and form $w_i = 1\eta_i + (\partial\theta)z_i$ for $i = 1, \ldots, N$. Use the results to form an empirical estimate of G^θ_*
3. Use these to form an empirical estimate of $v_t(\theta) = G^{\theta-}_*(u_t)$.
4. Compute
$$\epsilon_t(\theta) \equiv \frac{v_{2t}(\theta) - v_{1t}(\theta)}{\theta} \ , \qquad t = 1, \ldots, n \ .$$

5. Compute the log–likelihood

$$\ell(\theta) = \sum_{t=1}^{n} \epsilon_t^2(\theta)$$

6. Choose θ such that $\ell(\theta)$ is minimum

Here are some assertions I think are true about the meta–gamma.

- The means on the η_i or ϵ_i should not make a difference since we are transforming to percentiles. Or maybe it is just the relative means or something.
- By tuning the gamma's we can pretty much describe any chi-squared/normal behaviour
- The sum of independent gamma's is gamma so we can easily figure out G_ψ etc.
- If $\psi \to 0$ then the components of y approach perfect dependence.
- If $\psi \to \pm\infty$ then the components of y approach independence.

34 Weibull losses

Suppose the Weibull distribution function

$$u = F(y) = 1 - \exp\left\{-\left(\frac{y}{\alpha}\right)^{\beta}\right\} \qquad \Rightarrow \qquad F^{-}(u) = \alpha\left\{-\ln(1-u)\right\}^{1/\beta} \ .$$

indicating that if u is uniform then the right hand transformation is Weibull. With this specification

$$s(u) = \sum \alpha\left\{-\ln(1-u)\right\}^{\frac{1}{\beta}} \ ,$$

Further note the density of corresponding to $F(y)$ is

$$f(y) = \frac{\beta}{\alpha}\exp\left\{-\left(\frac{y}{\alpha}\right)^{\beta}\right\}\left(\frac{y}{\alpha}\right)^{\beta-1} = \frac{\beta}{\alpha}\left(\frac{y}{\alpha}\right)^{\beta-1}\{1-F(y)\}.$$

implying the product out the front is the hazard.

Further.

$$\mathrm{d}s = \sum \frac{\mathrm{d}u}{f\{F^{-}(u)\}} = \sum \frac{\alpha\left\{-\ln(1-u)\right\}^{\frac{1}{\beta}-1}}{\beta(1-u)}\ \mathrm{d}u \ .$$

which defines the iso–loss curves.

In practice the parameters α and β are estimated for each marginal distribution $F_i(y_i)$. The mean and variance associated with the marginal distribution i are then

$$\alpha\Gamma\left(1+\frac{1}{\beta}\right) \ , \qquad \alpha^2\left\{\Gamma\left(1+\frac{2}{\beta}\right)-\Gamma^2\left(1+\frac{1}{\beta}\right)\right\}$$

35 Open question

We note that $s(u) = \sum F_i^-(u_i)$ for some appropriate functions ϕ_i. The question is whether we can structure things such that this sum is nice. One procedure is where we first choose u_1 uniformly and compute $F_1^-(u_1)$. Next choose $u_2 \sim u_2|u_1$ and compute $F_2^-(u_2)$. Or we can just do the Gibbs sampling trick where we take $u_1|u_2$ then the other way round.

36 Logistic losses

Suppose

$$s(u) = \sum \frac{\alpha u^\beta}{1-u^\beta} = \sum \frac{\alpha}{u^{-\beta}-1}$$

37 Iso–loss curves

38 Estimation in old notation

To estimate λ take the observations y and initially transform to the quantiles

$$u \equiv F_*(y) = G_\lambda(\chi_1^2 + \lambda\cdot\epsilon) \ .$$

Next, for given θ transform u as

$$G_\theta^-(u) = (G_\theta^- \circ G_\lambda)(\chi_1^2+\lambda\cdot\epsilon) \approx (G_\theta^- \circ G_\lambda)(\chi_1^2) + (G_\theta^- \circ G_\lambda)'(\chi_1^2)\lambda\cdot\epsilon \ .$$

If $\theta = \lambda$ then this reduces to $\chi_1^2 + \lambda\cdot\epsilon$ which has the known distribution. Hopefully things are invariant to the base category.

In this formulation λ is unknown. To estimate it proceed as follows:

If (**??**) holds $\hat{\epsilon}_t(\lambda) \sim N(0, I)$. So the aim is to find θ such that

Generalizations

More generally suppose

$$y = (F_*^- \circ G_\lambda)(X\chi_1^2 + H\epsilon) \ , \qquad \epsilon \sim N(0, I) \tag{17}$$

To estimate we

Bivariate copulas

In the following X and Y are scalar random variables. The aim is to synthesize a bivariate copula with required properties.

1. Consider
$$(G_{X+Y} \circ G_X^-)(u) \equiv G_{X+Y}\{G_X^-(u)\} \ .$$
This is the quantile in the $X + Y$ distribution if X is at quantile u in the X distribution and $Y = 0$. Thus if $G_Y(0) = 0$ (ie $Y \geq 0$) it is the minimum quantile of $X + Y$ if X is at quantile u of the X distribution.

2. Suppose X and Y are independent chi–squareds each with 1 degree of freedom. Then $X + Y$ is chi–squared with 2 degrees of freedom. Hence $(G_{X+Y} \circ G_X^-)(u)$ is the chi-squared 2 degrees of freedom distribution evaluated at quantile u of the chi-squared 1 degree of freedom distribution.

3. Suppose $X \sim \chi_1^2$ and $Y \sim N(0, \sigma^2)$. Then $(G_{X+Y} \circ G_X^-)(u)$ is the $\chi_1^2 + N(0, \sigma^2)$ distribution evaluated at quantile u of the χ_1^2 distribution.

4. Consider $G_{X+Y}\{G_X^-(u) + G_Y^-(v)\}$. This the quantile of $X + Y$ when X is at quantile u and Y is at quantile v.

5. Consider $G_{X+Y}\{G_X^-(u) + Y\}$. This is a random variable – the quantile in the $X+Y$ distribution when X is at quantile u of the X distribution. Further $\mathrm{E}\left[G_{X+Y}\{G_X^-(u) + Y\}\right]$ and $Var\left[G_{X+Y}\{G_X^-(u) + Y\}\right]$ are the expected value and variance of the quantile in the $X + Y$ distribution if X is at quantile u in the X distribution, respectively.

6. Suppose $X \sim \chi_1^2$ and $Y \sim N(0, \sigma^2)$. Then
$$\mathrm{E}[G_{X+Y}\{G_X^-(u) + Y\}] = u \ , \qquad Var[G_{X+Y}\{G_X^-(u) + Y\}] =$$
To prove the first relation, note that
$$P\{G_X^-(u) + Y < y\} = P(Y < y - G_X^-(u)\} = G_Y\{y - G_X^-(u)\} \ .$$
that is a shifted version of G_Y.

Runoff triangles and copulas

Let L_k be the liability distribution associated with runoff triangle k. Suppose the joint distribution is modelled with a copula $C(u, v)$. Thus the joint distribution of (X, Y) is

$$y \sim C \circ F_*$$

Want $\iota' y$. Let's think of things in two dimensions:

1. Draw $u \sim C$ and
2. Put $y = F_*^- u$
3. Determine $y \sim F_*^- u$
4. AHP for determining a copula
   (a) For each u give probability distribution of v
   (b) Draw $u_1 \sim U$, $u_2 \sim C_{2|1}$
   (c) $y_1 = F_1^- u_1$, $y_2 = F_2^- u_2$.

AHP with copulas

Suppose $0 \leq u \leq 1$ is given and let U_1, U_2 and U_3 be risk quantiles. Suppose we know $U_1 > u$. We now ask how much more likely this makes $U_2 > u$ and $U_3 > u$. Similarly if $U_2 > u$ how much more likely does this make $U_1 > u$ and $U_3 > u$? In tabular format:

If $U_1 > u$ what is the probability of:

$P(U_1 > u)$	0.5	.25	0.10	0.05	0.025
$P(U_2 > u \mid U_1 > u)$	1	1	1	1	1
$P(U_3 > u \mid U_1 > u)$	0.5	0.25	0.10	0.05	0.025

This table indicates U_2 is perfectly dependent while U_3 is independent. Since

$$P(U_2 > u | U_1 > u) = \frac{P(U_2 > u, U_1 > u)}{P(U_1 > u)} = \frac{P(U_1 > u | U_2 > u) P(U_2 > u)}{P(U_1 > u)} = P(U_1 > u | U_2 > u)\,,$$

it makes no further sense to ask the question the other way round.

Conditional Probability

We have a risk vector y. Define $u = F_*(y)$. Then the marginal distributions of u are uniform. Next choose u_1. We then have an effect on the conditional distributions of the remainder. Hence construct $F_*(u|u_1)$ which again have

uniform marginals, albeit of one dimension less. The advantage is that if the effect of u_1 is simple on all the uniform marginals then we can keep track of everything and hopefully

Example copulas

38.0.1 Clayton

This depends on a single parameter δ and is defined as

$$C(u) \equiv \left(\Sigma u^{-\delta} - d + 1\right)^{-1/\delta}$$

It is the distribution of

$$u = \left\{1 - \frac{\ln(v)}{\gamma}\right\}^{-1/\delta}$$

where $\gamma \sim \Gamma(\delta^{-1}, 1)$ and $v \sim U^d(0,1)$. In the last step things are done componentwise.

38.0.2 Gumbel

This also depends on a single parameter δ and is defined as

$$C(u) \equiv \exp\left[\left\{-\Sigma(-\ln u)^{\delta}\right\}^{1/\delta}\right] , \qquad \delta \geq 1 .$$

It is the distribution of

$$u = \exp\left\{-\left(-\frac{\ln v}{s}\right)^{1/\delta}\right\} ,$$

where $s \sim \mathit{Stable}$ and $v \sim U^d(0,1)$.

Problem

Note that for a bivariate copula $f(v|u) = f(u,v)/f(u) = f(u,v) = f(u|v)$. Choosing $u \sim f(u)$ and then $v \sim f(v|u)$ produces a draw $(u,v) \sim f(u,v)$. But how do we constrain

So why does

One distribution is the beta distribution

$$f(v|u) = cv^{\alpha-1}(1-v)^{\beta-1} , \qquad \mathrm{E}(v|u) = \frac{\alpha}{\alpha+\beta} , \qquad \mathrm{cov}(v|u) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} ,$$

where $\alpha > 0$ and $\beta > 0$. Now we want to make α and β a function of u. A way of doing this is to put $\alpha - 1 = -a\ln(u)$ and $\beta - 1 = -b\ln(1-u)$ to yield

$$\ln\{f(v|u)\} = c - a\ln(u)\ln(v) - b\ln(1-u)\ln(1-v) = \ln\{f(u|v)\} \ .$$

where $a > 0$ and $b > 0$. Since $0 < u < 1$ then $\alpha > 1$ and $\beta > 1$.

Now $0 < x < 1$ and hence consider $\alpha = x/(1-x)$ implying $\alpha > 0$. Then $\alpha - 1 = (2x-1)/(1-x)$. Note when $x = 1/2$ then $\alpha - 1 = 0$.

So suppose $\beta = 1$

$$f(y|x) = cy^{(2x-1)/(1-x)} \ .$$

When $x = 0.5$, $f(y|x) = 1$. As

Point is to relate α and/or β to x, the conditioning variable. When x increases we want the mean to increase and vice versa. So maybe make $\alpha = x/(1-x)$, $\beta = 1$ in which case $\alpha - 1 = 1/(x-1)$ and

$$f(y|x) = cy^{1/(x-1)}$$

However further useful information is provided by $P(U_3 > u|U_2 > u)$. Natural questions that arise are:

- Is the fact that we are conditioning both on u important?
- What are the critical u points to be considered?
- How do we measure consistency of choice across the different combinations?
- If we do have tables as above how do we combine them?

Combining tables.
Sampling scheme:

- Select random U_1 from the uniform
- Find max table u such that $U_1 > u$ in the table
- With the table probability select U_2

Thus if it makes it x_u times more likely then since $P(U_2 > u) = 1 - u$ then

$$P(U_2 > u|U_1 > u) = x_u(1-u) \ , \qquad P(U_2 \leq u|U_1 > u) = 1 - x_u(1-u) \ .$$

Thus

$$\frac{P(U_2 > u|U_1 > u)}{P(U_2 \leq u|U_1 > u)} = \frac{x_u(1-u)}{1 - x_u(1-u)} \ .$$

Mortality smoothing with copulas

Suppose Y is a matrix of mortality measurements with rows indicating age i and columns time t. Define column vectors y_t and y_i as column t and row i of Y, respectively. There are two ways of smoothing the matrix Y:

38.0.3 Age copula

The copula C models correlation across age and $y_t = F_*^- u_t$ where $u_t \sim C$. Age specific behavior across time defines the components of F_* and are estimated from the ergodic average $\hat{F}_*(s) = \mathcal{E}[y_t < s]$. Special cases are:

1. Time profiles the same for each age: $F_* = 1 \odot h$ where h is a scalar function and estimated as $\hat{h}(s) \equiv \mathcal{E}\{\hat{F}_*(s)\} = \mathcal{E}\{\mathcal{E}[y_t < s]\}$ where the outer expectation is the ensemble, across age, average.

2. Perfect dependence across age. This implies all the component time series in y_t are driven by a single source of noise: $u_t = 1v_t$ with $v_t \sim U$.

3. Independence across age implies $u_t \sim \Pi$. This model invokes the unlikely assumption that each age moves independently over time.

4. The meta–Gaussian model correlation matrix R models the contemporaneous correlation between ages with $\hat{R} = \mathcal{E}(z_t z_t')$ where the z_t are the normits $z_t \equiv (\Phi_*^- \circ F_*)y_t \sim \Phi^R$. Generally the correlation between close ages is higher. An extreme case is where the correlation between ages i and j is $\rho^{|i-j|}$ for $0 \leq \rho \leq 1$. The case $\rho = 1$ is equivalent to 2 while $\rho = 0$ to 3 above.

38.0.4 Serial copula

The copula C models serial correlation and $y_i = F_*^- u_i$ where draws $u_i \sim C$ correspond to different ages. Time specific mortality curves are the components in F_* and are estimated from the ensemble (age profile) average $\hat{F}_*(j) \equiv \mathcal{E}[y_i < j]$. Special cases are:

1. Age profiles are the same across time. This implies each marginal in F_* is the same: $F_* = 1 \odot h$ where h is univariate and $(1 \odot h)y_i$ applies h componentwise. Thus $y_i = (1 \odot h)^- u_i = (1 \odot h^-)u_i$. Function h is estimated from the ergodic average of the empirical ensemble (age profile) distributions $\hat{h}(j) \equiv \mathcal{E}\{\hat{F}_*(j)\} = \mathcal{E}\{\mathcal{E}[y_i < j]\}$ where the outer $\mathcal{E}$ works across the components of the vector.

2. Perfect serial dependence implies $u_i = 1v_i$ where $v_i \sim U$. Combined with a single age profile yields $y_i = (1 \odot h^-)1v_i = 1h^- v_i$.

3. Matrix R in the meta–Gaussian model is a serial covariance matrix, estimated from the normits $z_i \equiv (\Phi_*^- \circ F_*)y_i \sim \Phi^R$ as $\hat{R} = \mathcal{E}(z_i z_i')$. With a fixed mortality curve $z_i = \{\Phi_*^- \circ (1 \odot h)\}y_i$ implying each component of y_i is transformed with a fixed h.

4. With the meta–Gaussian model the stationary AR(1) case is where entry (t, s) of R is $\rho^{|t-s|}$ with $0 \leq \rho \leq 1$. Estimate $\hat{\rho}$ is derived from $\hat{R}$ by averaging the first off diagonal.

Which copula modelling is likely to be more appropriate? Copulas are preferred for variables which have distinct marginal behavious.

38.0.5 Modelling age at death

Let $F_t(i)$ be the age–at–death distributions at time t and from this define F_*. Thus F_* is a vector of marginals corresponding to different points of time. Corresponding to $y_i = F_*^- u_i$ with $u_i \sim C$ write, for each age i, u_i as the vector with components $F_t(i)$. A copula C serves to model dependence across age.

The above approach serves to define u_i from the physically constructed age at death distribution, not actual deaths. Of course the physical age–at–death distribution

38.0.6 Connection to usual studies

The distribution of age–at-death in a given year is

$$F(y) = P(Y < y) = \sum_x P(Y < y|X = x)P(X = x) = \int_x F(y|x)f(x)$$

Differentiating with respect to y implies the density is

$$f(y) = \int_x f(y|x)f(x)$$

$$F(y|x)F_X(x)] - \int_x f(y|x)F_X(x)$$

38.0.7 Mortality smoothing across time

Time dependence is introduced by assuming

$$N_* y_t = Z\alpha_t + G\epsilon_t \ , \qquad \alpha_{t+1} = T\alpha_t + H\epsilon_t \ .$$

where $N_* \equiv \Phi_*^- \circ F_*$. Hence the steps in this process are that we first construct the normits $N_* y_t$ from each of the ages over time. We then fit a standard state space model to these normits. But there must be a cycle here since the normits are estimated over time. Given a forecast $\hat{z}_t$ of the normits the forecast of y_t is $\hat{y}_t = N_*^- \hat{z}_t$ where $N_*^- \equiv (\Phi_*^- \circ F_*)^- = F_*^- \circ \Phi_*$

An alternative is to define $N_+ = \Phi_*^- \circ F_+$ where F_+ computes the quantiles of all the available data considered as one group. Thus the measurement equation becomes $N_+ y_t = Z\alpha_t + G\epsilon_t$.

A very simple model is

$$N_* y_t = 1\alpha_t + G\epsilon_t \ , \qquad \alpha_{t+1} = \rho\alpha_t + \eta_t \ .$$

BvN

Theorem 1: The Birkhoff - von Neumann Theorem [Marcus (1960), pg. 215] A doubly stochastic matrix A can be expressed as , where are nonnegative real numbers summing to 1 and are permutation matrices.

Therefore any copula can be discretized and represented in matrix form based on the partition into regions of equal size. Then, this matrix can be decomposed into linear combination of permutation matrices which can enhance the understanding of dependence structure and help to form a basis for multivariate data analysis. In the following sections we will set up BvN decomposition and copula connection, study the dependence bones and skeletons, discuss implications of the decomposition and generating copulas, and propose an approach to decompose copulas.

Power copulas

Suppose C is a copula. For positive integer p consider C^p. This is a distribution on the unit hypercube. The marginals are not uniform. However we can construct the copula

$$C^{(p)} \equiv C^p \circ C_*^{p-}$$

What is the dependence structure here? The procedure is thus:

- Choose a given copula C. For example $C = \Phi^R \circ \Phi_*^-$

- Power it. For example $C^2 = (\Phi^R \circ \Phi_*^-)^2$. Note powering it means take the appropriate pointwise power not taking the function twice.
- Turn the powered version into a copula.
- Use the copula to model dependence.

Tail correlation

We want an effective method to construct the correlation in the tail of a distribution. For example when dealing with the Gaussian distribution there is a single correlation that describes the strength of the relationship no matter where you are in the distribution. How can we model the situation where there is more correlation in the tails? Here is a try. Think of the bivariate situation. Divide the unit square into smaller squares. In each square we correlate the normits of the quantiles. Thus in each little square we have a bivariate normal and we compute the correlation. To get the overall correlation we correlate the correlations.

Is this effective? One way to think about things is to think about

$$\tau(x, y) \equiv P(X > x | Y > y) = \frac{P(X > x, Y > y)}{P(S > y)} = \frac{S(x, y)}{S(y)}$$

What is a smooth way of perturbing the normal to get more emphasis in the tail. One way of doing this is to think of the distribution on each margin and then multiply by another distribution

$$G \circ F$$

Call G the "tail emphasizer." In the extreme for example we can have G as a step function with knot κ. Then result is a distribution which places a discrete mass at the knot and then as usual. What if we take a ramp function...

Note that a copula is a function of many variables and returns a probability. So you can't apply a copula repeatedly. What happens if you take a copula G and and generate a second copula from it. That is you construct $C_C \equiv C \circ C_*$. But the marginals C_* of C are uniform. Hence $C_C = C \circ I = C$

What about taking a Gaussian copula and applying it repeatedly? For example $G \equiv \Phi^R \circ \Phi_*^-$ is the Gaussian copula with correlation matrix R. Using What about G^p. Note that G is a copula. Hence it takes probabilities to probabilities. Thus G^p is a copula for any integer p. Note $G^0 = I$, the uniform copula.

Survival copulas

Copulas are typically thought of as the joint distribution on the square. Alternatively let S denote a survival function:

$$S(y) = P(Y > y)$$

We can then think of a survival copula on the square. Thus a survival copula is a survival function on the unit cube with uniform marginals. Note for positive random variables $S(0) = 1$ and $S(y) \to 1$ as $y \to \infty$.

Suppose $y \sim S$, indicating y is a random vector with survival function S. Consider the "survival" transform

$$P\{S(Y) > y\} = P\{Y < S^-(y)\} = 1 - S\{S^-(y)\} = 1 - y$$

Thus let D be a survival copula on the unit hypercube with uniform marginals and S_* a vector of survival marginals and define

$$S_D \equiv D \circ S_* \ , \qquad D_S \equiv S \circ S_*^-$$

Total liability distribution

Consider

$$\phi(s) \equiv P(\iota' F_*^- u \leq s) \ , \qquad s_\alpha = \phi^-(\alpha) \ ,$$

where $u \sim C$ and where $0 \leq \alpha \leq 1$ is given. What is a good way of estimating $F(s)$? Further what is a good way of determining s_α given α? For example we may wish to estimate s_α, corresponding to $\alpha = 0.75$, the 75th percentile.

One approach is to repeatedly sample $u \sim C$ and hence $\iota' F_*^- u$. These realizations are transformed to quantiles and in turn the quantiles are used to define s_α.

If we are interested in say the α quantile then hopefully we need only consider u in the triangular pyramid of points lying above

Also important is the copula to be used. What is a good copula?

Can we simplify. For this purpose look at the 2D case and consider the point (u, v) and suppose the quantile associated with this point insofar as the sum is concerned is α. Suppose we move x to $x - \delta$ where $\delta > 0$. At the same time we move y to $y + \delta$ and hence the sum stays the same. The change in quantile in the x distribution is from $u = F_X(x)$ to $F_X(x - \delta) \approx F_X(x) - \delta f_X(x) = u - \delta f_X(x)$. That is the change in the quantile is approximately $-\delta f_X(x)$. Similarly the change in the y quantile is approximately $\delta f_Y(y)$. Thus the relative change in the y quantile to that in the x quantile is

$$-\frac{f_Y(y)}{f_X(x)} \ .$$

Fix $s > 0$ and consider $\Omega_s = \{(x, y) : x + y \leq s\}$. Clearly $(s, 0) \in \Omega_s$. Want $F_*(\Omega_s)$, the quantile set corresponding to Ω_s. Note

$$F_*(\Omega_s) = \{u : \iota' F_*^- u \leq s\} \ .$$

What does this set look like? Here we look at the 2D situation. First consider (in 2D) $(u_s, v_s) = F_*^-(s, s)$. These are the intercepts points, ie the quantiles in the marginal distributions that yield s. Hence $(u_s, 0)$ and $(0, v_s)$ are in $F_*(\Omega_s)$. Moving from $(0, s)$ to $(\delta, s - \delta)$ keeps us in Ω_s. The corresponding movements in (u, v) space is from $(0, v_s)$ to $(\delta f_X(0), v_s - \delta f_Y(s))$. Thus the slope is $-f_Y(s)/f_X(0)$.

The above argument can be generalized to any point when moving from (u, v), presumed to be such that the corresponding $x + y = s$. The change in v relative to a change in u is $-f_Y(x - s)/f_X(x)$. Hence the boundary is defined by the differential equation

$$\partial v = \frac{-f_Y\{F_X^-(u) - s\}}{f_X\{F_X^-(u)\}} \partial u \ ,$$

with starting condition $(0, F_Y^-(s))$.

Once we have the boundary then we sample repeatedly from the copula and check if the sampled point lies on the right side of the boundary. If so we count 1 and otherwise 0 and averaging we determine the quantile of s. Alternatively we determine say the α quantile in F_Y called y_α say. We then use the above differential equation to define the boundary using y_α in place of s. We repeatedly sample from the copula and check if we are on the right side of the boundary and count 1 or 0. Averaging the 1's leads to the quantile of y_α. Generally expect that the quantile of y_α will be much bigger than α. Interesting to see how much bigger it is.

Here is an interesting graph. Compute $y_\alpha \equiv F_*^- 1\alpha$. Then compute the sum quantile from each component of y_α that is F Presumably the sum quantiles in each case are much less than α.

Alternatively, we simply sample from the copula and just construct

$$s = \iota' F_*^- u \ , \qquad u \sim C \ .$$

and compute the quantiles associated with the generated sample.

Suppose $F_X(s) = \alpha$. Thus the point $(s, 0)$ in the set of $x + y \geq s$. Decreasing x by a small amount and increasing y by the corresponding amount so as to keep the sum fixed implies the ratio of the percentiles changes by

$$-\frac{f_Y(0)}{f_X(s)} \ .$$

It is not unreasonable to suppose $f_Y(0) = 0$. Hence the change in the percentile... Hence provided $f_Y(0) < f_X(s)$ the change in the percentile Suppose $F_X(x_\alpha) = \alpha$. Thus x_α is the α quantile in the X distribution. Thus $(x_\alpha, 0)$ is in the set of $x + y \geq s_\alpha$.

we wish to determine s_α say such that In two dimensions this defines a line, in three dimensions a surface, and so on. The hypersurface defines a region with a probability

$$P\{\iota' F_*^- u \leq s\} =$$

Consider the two dimensional copula $C(u, v)$. Consider the "line" $v_s(u)$ such that $\iota' F_*^-\{u, v_s(u)\} = s$. Thus for each $0 \leq u \leq 1$, $v_s(u)$ is that point... Note that $v_s(0)$ is the quantile of s in the Y distribution, while $v_s(1)$ is the quantile of $s - \max(X)$ in the X distribution.

Alternatively consider any point u. Then this maps to point say $y = F_*^- u$ and sum $\iota' y = \iota' F_*^- u$. Further the copula $C(u) = P(U < u)$. Thus $P(\iota' F_*^- U < s) =$

Tail copulas

In liability forecasting only the upper tails of distributions are of interest. This motivates the use of tail copulas. A tail copula is a copula that joins marginal distributions in the tails. In particular define F_α such that

$$F_\alpha(y) = \begin{cases} (1-\alpha)^{-1}(F_* - \alpha)(y) \ , & F_*(y) \geq \alpha \ , \\ 0 \ , & F_*(y) < \alpha \ . \end{cases}$$

Thus F_α is the vector of tail marginal distributions. We now copulate the tail marginals as $C \circ F_\alpha$ which is a distribution of

Consider a bivariate copula $C(u, v)$. This yields tail probabilities

$$C^\alpha(u, v) = C(u, v | u > \alpha, v > \alpha) = \frac{C(u, v)}{1 - C(0.5, 0.5) + ...}$$

say. This is a joint distribution on the upper right hand square. We can turn this joint distribution into a copula viz.

$$C^\alpha \circ C_*^\alpha \qquad F_{C^\alpha} \equiv C^\alpha \circ F_*$$

Sampling from distributions generated with a copula

To simulate from $C \circ F_*$, proceed as follows:

1. Decompose C into $C = C_1 C_{2|1} C_{3|1,2} \cdots$

2. Draw $u_1 \sim C_1 \equiv U$, $u_2 \sim C_{2|1}$, $u_3 \sim C_{3|1,2}$ and so on leading to $u \equiv (u_1, u_2, u_3, \ldots,)'$.

3. Put $y = F_*^- u$.

Then $y \sim C \circ F_*$. The potential difficulty is finding/specifying the conditional distributions. What are reasonable specifications. How do we determine what a reasonable specification is. Maybe, similar to time series, we should forget about specifying a global copula but concentrate on just specifying a sequence of conditional copulas starting from $C_1 = U$. Note that

$$C(u_2|u_1) \equiv \frac{C(u_1, u_2)}{U(u_1)} = \frac{C(u_1|u_2)u_2}{u_1} \qquad \Rightarrow \qquad \frac{C(u_2|u_1)}{C(u_1|u_2)} = \frac{u_2}{u_1} \ .$$

Hence $C(u_2|u_1) \geq C(u_1, u_2)$.

Suppose $u_t \sim C$ and $u_t = (v_t', w_t')'$

Tail dependence

For two random variables X and Y consider

measure	definition	X and Y independent	X and Y perfectly dependent
$\phi(u)$	$\mathrm{E}\{[F(Y) > u]\|F(X) > u\}$	$1-u$	1
$\psi(u)$	$\mathrm{E}\{F(Y)\|F(X) > u\}$	$1/2$	$(1+u)/2$
$\rho(u)$	$\mathrm{cor}\{F(X), F(Y)\|F(X) > u\}$	0	1

Where $[F(Y) > u]$ is 1 if $F(Y) > u$ and 0 otherwise. Thus measures indicating the departure from independence are $\phi(u)/(1-u)$ and $2\psi(u)$ while measures indicating the departure from perfect dependence are $\phi(u)$ and $2\psi(u)/(1+u)$.

Since

$$\phi(u) = \frac{P\{F(Y) > u, F(X) > u\}}{P\{F(X) > u\}} = \frac{P\{F(Y) > u, F(X) > u\}}{1-u} \ .$$

Hence $\phi(u)$ is symmetric in X and Y. This is not a property enjoyed by the other two measures of tail dependence.

Note $2\psi(u) - \phi(u) = u$ for both perfectly dependent and independent situations. Alternatively $2\psi(u) + \phi(u) = 2 - u$ if independent and $2 + u$ if perfectly dependent.

A further variant is the moment generating type function

$$\mathrm{E}\{\mathrm{e}^{sF(Y)}|F(X) > u\} \ ,$$

For each u, the conditional probability is estimated as the empirical proportion of the quantiles of Y that exceed u given the empirical quantile of X exceeds u. As u approaches 1, $\phi(u)$ may approach some number bigger than 0.

Let's define the normits as

$$x$$

By Bayes' theorem

$$\phi(u) = \frac{P\{F(X) > u|F(Y) > u\}}{P\{F(X) > u\}} P\{F(Y) > u\} \ .$$

Hence

$$\phi_{Y|X}(u) = \phi_{X|Y}(u)$$

Note that $\phi(u) \equiv 1$ if Y and X are perfectly dependent while $\phi(u) = 1 - u$ under independence. Hence $\phi(u)$ measures the percentage reduction in dependence while $\phi(u)/(1-u)$ measures how far things are away from independence.

Note that

$$P\{F(X) > u, F(Y) > u\} = P\{F(X) > u\}P\{F(Y) > u|F(X) > u\} = (1-u)\phi(u) \ .$$

This implies

Similarly we can define

$$\rho(u) \equiv \mathrm{cor}\{F(X), F(Y)|F(X) > u\} \ .$$

If X and Y are independent $\psi(u) = 0.5$ and $\rho(u) = 0$. Positive tail dependence is $\psi'(u) > 0$.

Note that

$$0.5 = \psi(u)(1-u)+\mathrm{E}\{F(Y)|F(X) < u\}u \qquad \Rightarrow \qquad E\{F(Y)|F(X) < u\} = \frac{0.5 - \psi(u)(1-u)}{u}$$

When there is perfect dependence between Y and X then $\psi(u) = (1 + u)/2$. Hence a reasonable measure of tail dependence is

$$\frac{2\psi(u)}{1+u}$$

If there is perfect tail dependence then this should be close to 1. If there is no tail dependence then $\psi(u) = 1/2$ and this measure is $1/(1+u)$ which approaches 1/2 as

When there is dependence between Y and X then $\mathrm{E}\{F(Y)|F(X) > u\}$ will increase with u.

Theoretically the expression is

$$= \frac{P\{F(X) > u, F(Y) > u\}}{P\{F(Y) > u\}}$$

$$= \lim_{u\uparrow 1} \frac{1 - P\{F(X) < u\} - P\{F(Y) < u\} + P\{F(X) < u, F(Y) < u\}}{1 - P\{F(Y) < u\}} = \lim_{u\uparrow 1} \frac{1 - 2u + C(u,u)}{1-u} .$$

An empirical estimate is to take the proportion of cases where all quantiles exceed u of the proportion of cases For an empirical estimate we do the following:

1. For each quantile u determine the proportion of cases where both variables have a bigger quantile

2. Divide by the proportion proportion of x variables have bigger quantile

Simulating using a Gaussian copula

Suppose the correlation matrix R is given and $z \sim \Phi^R$ then $\Phi_*(z) \sim \Phi^R \circ \Phi_*^-$. To prove this consider

$$P\{\Phi_*(z) < u\} = P\{z < \Phi_*^-(u)\} = \Phi^R\{\Phi_*^-(u)\} = (\Phi^R \circ \Phi_*^-)(u) ,$$

as required.

More generally suppose Ψ

Suppose we take $z \sim \Phi^R$ and consider z^2. Each component has a chi–squared distribution and hence $u = \chi_*^-(z^2)$ has uniform marginals.

Copulas for runoff triangles

Suppose x

38.0.8 Measurement noise

In many cases it is appropriate to add "measurement" noise. For example

$$y_t = F_*^- u_t + F_*^- \epsilon_t , \qquad u_t \sim C , \qquad \epsilon_t \sim \Pi .$$

In the case of the Gaussian copula a signal plus noise model is

$$y_t = (F_*^- \circ \Phi_*)(\sqrt{R - \sigma^2 I}\zeta_t + \sigma\xi_t) \ , \qquad (\zeta_t', \xi_t')' \sim \Phi^I \ .$$

Thus the covariance matrix of the overall disturbance is R. In this formulation σ^2 cannot exceed the smallest eigenvalue of R. A further refinement is where the copula from which u_t is drawn is a conditional copula.

References

[HarveyHarvey1989] Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

[JoeJoe1997] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall/CRC.

[Lawley and MaxwellLawley and Maxwell1971] Lawley, D. N. and A. Maxwell (1971). *Factor analysis as a statistical method*. Butterworths.

[McNeil, Frey, Embrechts, et al.McNeil et al.2005] McNeil, A., R. Frey, P. Embrechts, et al. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

[NelsenNelsen2006] Nelsen, R. (2006). *An Introduction to Copulas*. Springer.