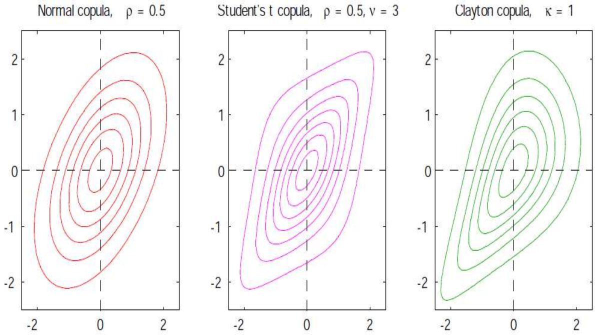
Dependence Measures and Extreme Value Theory

Leopoldo Catania
Aarhus University and CREATES
leopoldo.catania@econ.au.dk

Examples

A variety of bivariate distributions with Normal margins



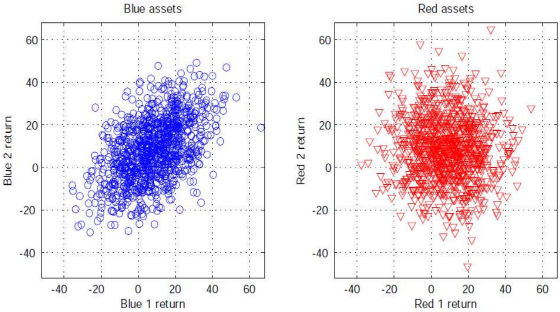
Why do we care so much about joint distribution

- One important economic application where the form of dependence matters is portfolio decision making.

- Consider the following illustration:
  - Two pairs of assets.
  - All assets are individually N(8%, $15\%^2$)
  - We will vary their dependence structure (copula) and consider the outcome.

Example 1

Which pair of assets would you prefer?



Example 1

Consider an equally weighted portfolio:
Blue Assets ($\rho = 0.5$):

$$
\begin{aligned}
E[\frac{1}{2}X + \frac{1}{2}Y] &= 8\% \\
Var[\frac{1}{2}X + \frac{1}{2}Y] &\approx 13^2\%
\end{aligned}
$$

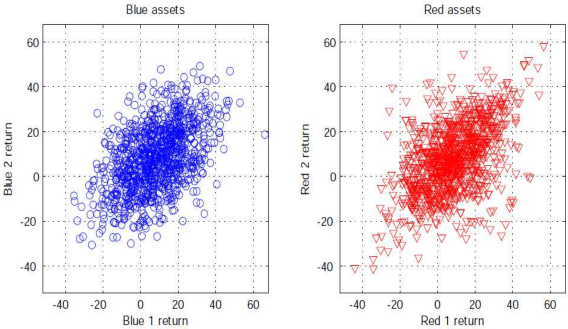
Red Assets ($\rho = 0$):

$$
\begin{aligned}
E[\frac{1}{2}X + \frac{1}{2}Y] &= 8\% \\
Var[\frac{1}{2}X + \frac{1}{2}Y] &\approx 10.5^2\%
\end{aligned}
$$

We should prefer the red assets.

Example 2

Both have N(8,15^2) margins and $\rho = 0.5$. Which pair of assets would you prefer?



Example 2

All assets have the same marginal distributions, and correlation of both pairs is 0.5.

- So mean-variance comparisons of portfolios will not distinguish between the two pairs.
- But we could see that the red pair (which had a Student's t copula) had more joint crashes and booms than the blue pair (Normal copula).
- This leads to more kurtosis in a portfolio of the red assets.

Summary

These simple examples show that pairs of variables with the same marginal distributions and the same degree of (linear) correlation can be consistent with many different bivariate distributions.

- These joint distributions differ, broadly, in:
  1. The degree to which large events are correlated (tail dependence).
  2. The degree to which negative events have different correlation to positive events (asymmetric dependence)
- In portfolio applications, we know that risk-averse investors have clear preferences over these dependence structures
- To study these dependence structures we need more flexible models of dependence, and richer measures of dependence.

Pearson's Correlation

- The most widely-used measure of dependence is Pearson's linear correlation:

$$Corr[Y, Z] = \frac{Cov[Y, Z]}{\sqrt{V[Y]V[Z]}}$$

- This simple measure contains a lot of information, but it suffers from some limitations.

Limitations of Pearson's correlation

- Many different dependence structures (copulas) can have identical linear correlation coefficients (as we saw in the portfolio illustration).

- For example, when combined with N (0, 1) margins, the following copulas all lead to linear correlation of 0.5:
  - Normal with $\rho = 0.5$
  - Student's with $\rho = 0.5$
  - Clayton with $\theta = 1$
  - Gumbel with $\theta = 1.5$

Limitations of Pearson's correlation

- The range of linear correlation is affected by the marginal distributions.
- The actual range of possible values for linear correlation is narrower, and depends on the marginal distributions.
- If the margins are identical (up to affine transformations) then the full range is possible $[-1, 1]$
- If the margins are different then the actual range of possible values can be much narrower.

An Extreme Example

Perfectly dependent variables can have linear correlation arbitrarily close to zero (rare, but possbile).

- As in Embrechts, et al. (2002), consider two perfectly dependent variables:

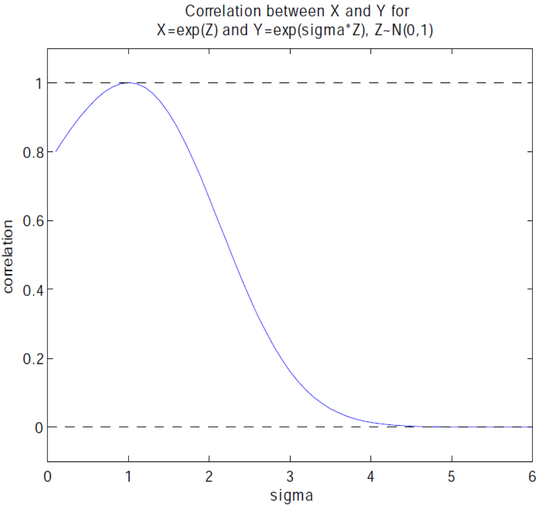
$$\begin{aligned} Z &\sim N(0,1) \\ X &= exp(Z) \\ Y &= exp(\sigma Z) \end{aligned}$$

then it is possible to show that

$$Corr[X,Y] = \frac{exp(\sigma) - 1}{\sqrt{\exp(1) - 1}\sqrt{\exp(\sigma^2) - 1}} \to 0$$

as $\sigma \to \infty$.

Limitations of linear correlation



Alternative Measures of Dependence: Spearman's Rank Correlation

Rank correlation is simply the linear correlation of the ranks of the variables.

- The smallest observation has a rank of 1, second-smallest a rank of 2, etc.
- Let $R^X =$ Rank of X in $(X_1, ..., X_T)$;
- For example, if $X = [5, 6, 1, 8, 3]$, the rank is

$$R^X = [3, 4, 1, 5, 2]$$

then, Spearman's rank correlation between X and Y is

$$\varsigma^{X,Y} = corr(R^X, R^Y)$$

Properties of Spearman's Rank Correlation

- The ranks of observations in a sample are unaffected by strictly increasing transformations, and so rank correlation is unaffected by such transformation.
- Consider the PIT, $U_i = F_i(X_i)$, then we have that :

$$\varsigma^{X,Y} = corr(R^X, R^Y) = corr(U_x, V_x)$$

- Thus rank correlation is purely a function of the copula of (X,Y).

Kendall's tau

- Kendall's tau is another widely-used measure of dependence. It is based on the proportion of *concordant* pairs of observations in a sample.
- Let x_t, y_t for $t = 1, ..., T$ be a sample of observations. There are $\frac{T!}{2(T-2)!}$ distinct pairs of observations (x_i, y_i) and (x_j, y_j).
- A pair of observations is concordant if $(x_i - x_j)(y_i - y_j) > 0$, else the pair is discordant
- Let c and d denote the number of concordant and discordant pairs of observations. Then

$$\tau = \frac{c - d}{c + d}$$

- The Kendall's tau is also function of the copula only

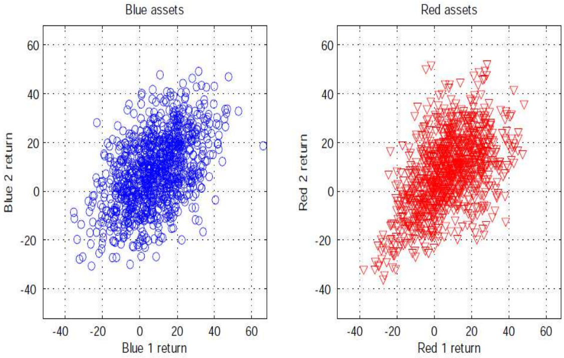
$$\tau = 4E[C(U_x, U_y)] - 1$$

Kendall's tau and Copulas

- The Kendall's tau of the Gaussian Copula is given by $\frac{2}{\pi}\arcsin(\rho)$
- The Kendall's tau of the t-Copula is given by $\frac{2}{\pi}\arcsin(\rho)$
- The Kendall's tau (τ) coefficient for the Gumbel copula is equal to $1-\frac{1}{\theta}$.
- The Kendall's tau of the Clayton copula is equal to $\frac{\theta}{\theta+2}$.

Example 3

Both have N(8,15^2) margins and $\rho = 0.5$. Which pair of assets would you prefer?



Example 3

All assets have the same marginal distributions, and correlation of both pairs is 0.5.

- Again, mean-variance comparisons of portfolios will not help here
- But we could see that the red pair (Clayton copula) had more joint crashes and fewer joint booms than the blue pair (Normal copula)
- This leads to more skewness and kurtosis in the portfolio of the red assets.

Asymptotic dependence

- The concepts or asymptotic dependence and independence, i.e. degree of association of tail events
- Important to remove the influence of marginal aspects (no effects on the asymptotic dependence).
- Given a pair of RV, X, Y, the analysis is carried out on $U \equiv F_X(X)$ and $V \equiv F_Y(Y)$.
- Specifically, we consider the behavior of $P(V > u|U > u)$.
- In case of perfect dependence, then $P(V > u|U > u) = 1$.
- In case of perfect independence, then $P(V > u|U > u) = P(V > u)$.

Quantile Dependence and Copulae

- The measure of quantile dependence are functions that are defined as

$$\begin{aligned}\lambda^L(q) &= P(U_x < q | U_y < q) = \frac{P(U_x < q, U_y < q)}{P(U_y < q)} \\ &= \frac{C(q, q)}{q}\end{aligned}$$

$$\begin{aligned}\lambda^U(q) &= P(U_x > q | U_y > q) = \frac{P(U_x > q, U_y > q)}{P(U_y > q)} \\ &= \frac{1 - 2q + C(q, q)}{1 - q}\end{aligned}$$

- The function $\lambda^L(q)$ can be interpreted as the probability that one variable lies in its lower q tail given that the other variable lies in its lower q tail.
- Different copulas imply different quantile dependences.

Tail Dependence

- Pushing the quantiles to the limit...
- The measure of negative tail dependence is defined as

$$\lambda^L = \lim_{q \to 0} P(U_x < q | U_y < q) = \lim_{q \to 0} \frac{C(q,q)}{q}$$

$$\lambda^U = \lim_{q \to 1} P(U_x > q | U_y > q) = \lim_{q \to 1} \frac{1 - 2q + C(q,q)}{1-q}$$

Tail Dependence and Copulae

- The Gaussian copula does not generate tail dependence for any $|\rho| < 1$.

- The t-copula has symmetric tail dependence if $\nu < \infty$

$$\lambda^U = \lambda^L = 2\mathcal{T}_{\nu+1}\left(-\sqrt{\nu+1}\sqrt{\frac{1-\rho}{1+\rho}}\right),$$

  where $\mathcal{T}_{\nu+1}(\cdot)$ is the cumulative distribution function of a univariate Student's t random variable with $\nu+1$ degrees of freedom.

- The degree of upper tail dependence for the Gumbel copula is equal to $2 - 2^{\frac{1}{\theta}}$, see Joe (1997).

- The Clayton copula implies a degree of tail dependence equal to $2^{(-1/\theta)}$.

Extreme Value Theory

- Modeling the entire distribution of the returns is useful but possibly complicated
- For some applications it is not needed to look at the entire distribution, e.g. VaR or expected shortfall.
- We can use specific tools created for the description of the tails.
- Extreme Value Theory is a set of statistical tools for the analysis of extreme realizations, both in univariate and in the multivariate context.
- The focus is on the so-called *tail index*.

Univariate Tail Estimation

Two approaches:

1. *Extrema approach*: distribution of the maxima/minima
2. *Tail approach*: exceedances over a given threshold.

- The extrema approach leads to the definition of a generalized extreme value distribution, *gev*.
- Given a sequence $(X_1, ..., X_T)$ of random variables, we define

$$M_T \equiv -min(-X_1, ..., -X_T) = max(X_1, ..., X_T)$$

- If $(X_1, ..., X_T)$ are *i.i.d.*, then

$$P(M_T \leq x) = [F_X(x)]^T$$

Univariate Tail Estimation

Theorem

Let X_t be a sequence of i.i.d. random variables. If there exists a location parameter $\mu_T \in \mathcal{R}$, and a scale parameter $\psi_T > 0$, and some non-degenerate cdf, H such that the limit distribution of the standardized extremes, $Y_T = \frac{M_T - \mu_T}{\psi_T}$ converges to H,

$$\lim_{T \to \infty} P\left(\frac{M_T - \mu_T}{\psi_T}\right) = H(y), \quad \forall y \in \mathcal{R}$$

then $F_X(\cdot)$ is said to belong to the domain of attraction of H, which is a standard extreme value distribution.

The gev encompasses the standard extreme value distributions,

$$H_\xi(y) = \begin{cases} exp(-(1+\xi y)^{-1/\xi}), & \text{if } \xi \neq 0, 1+\xi y > 0 \\ exp(-exp(-y)), & \text{if } \xi = 0 \end{cases}$$

where ξ is called tail index. If $\xi > 0$, we have the Frechet distribution, if $\xi = 0$ we have the Gumbel and if $\xi < 0$ we have the Weibull.

Quantile Plot

- The quantile plot is a nice graphical instrument to understand the limit distribution of the maxima of the returns.
- Consider $\tau = T/N$ sub-samples of size N over T periods.
- For each period compute $m_i = max(x_{(i-1)N+1}, ..., x_{iN})$.
- Order the maxima as, $\tilde{m}_1, ..., \tilde{m}_\tau$.
- For each $\tilde{m}_i$, compute the theoretical quantiles, $H^{-1}_{\xi,\mu,\psi}\left(\frac{1}{\tau+1}\right), ..., H^{-1}_{\xi,\mu,\psi}\left(\frac{\tau}{\tau+1}\right)$.
- Assume the distribution is Gumbel, then

$$H^{-1}_{\xi,\mu,\psi}\left(\frac{i}{\tau+1}\right) = -log\left(-log\left(\frac{i}{\tau+1}\right)\right)$$

- Plot $\tilde{m}_1, ..., \tilde{m}_\tau$ wrt $H^{-1}_{\xi,\mu,\psi}\left(\frac{1}{\tau+1}\right), ..., H^{-1}_{\xi,\mu,\psi}\left(\frac{\tau}{\tau+1}\right)$ on a 2-D plot.
- If the plot is concave, we have a Frechet distribution, if convex the limit distribution is Weibull.

ML estimation of the gev

- Consider an $i.i.d.$ sample, m_i for $i = 1, ..., \tau$.
- The log-likelihood function wrt $\theta = (\mu, \psi, \xi)$ is

$$L_\tau(\theta|m_i) = \sum_{i=1}^{\tau} \ell(\theta)$$

where

$$\ell_i(\theta) = -\log(\psi) - \left(\frac{1}{\xi} + 1\right) \log\left(1 + \xi \frac{m_i - \mu}{\psi}\right) - \left(1 + \xi \frac{m_i - \mu}{\psi}\right)^{-\frac{1}{\xi}}$$

- If $\xi = 0$,

$$\ell_i(\theta) = -\log(\psi) - \frac{m_i - \mu}{\psi} - \exp(-\frac{m_i - \mu}{\psi})$$

- The ML estimator has standard asymptotic for $\xi > -1/2$.

Tail approach

- The tail approach is based on modeling the tails of a distribution.
- We consider an iid sample, $(X_1, ..., X_T)$, with distribution $F_X(\cdot)$.
- Let u be a fixed real number, the threshold, then

$$F_u(y) = P(X_t - u \leq y | X_t > u) = \frac{F_X(y+u) - F_X(u)}{1 - F_X(u)}$$

  is the excess distribution function.
- The function

$$e(u) = E(X_t - u | X_t > u)$$

  is called mean-excess function.
- The excess distribution function can be approximated by the Generalized Pareto distribution (gdp)

$$F_u(y) \approx G_{\xi,\psi}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\psi} y\right)^{-1/\xi}, & \text{if } \xi \neq 0 \\ 1 - exp\left(-\frac{y}{\psi}\right), & \text{if } \xi = 0 \end{cases} \tag{1}$$

- The tail index of the gev is the same as the one of the gpd.

Hill estimator

- An estimate of ξ can be obtained also by a semi-parametric method.
- The Hill estimator is

$$\hat{\xi}^{H}_{(q,T)} = \frac{1}{q} \sum_{j=1}^{q} \log \left(\frac{x_{T-j+1,T}}{x_{T-q,T}} \right) \quad 1 \leq q < T$$

- Under the assumption that the distribution of X_t belongs to the Frechet, then

$$\sqrt{q} \left(\hat{\xi} - \xi \right) \rightarrow N(0, \xi^2) \tag{2}$$

- Bootstrap techniques for the selection of the optimal q have been proposed.

- Alternatively, an ML estimation can be performed (parametric setup).