Diversification benefit in Gaussian Aggregation Trees

JP Bruneton, jpbruneton@gmail.com
Formerly at SCOR
Department of mathematics, Namur University, Belgium

In collaboration with
F. Uffer, senior risk analyst at FinMod, SCOR
M. Dacorogna, Deputy CRO, SCOR

20/10/2011, ESSEC Business School, WG Risk

SCOR

Diversification benefit in Gaussian Aggregation Trees

Introduction (1)

Diversification…

- Diversification of risks is key to insurers and reinsurers, because
  - It lowers significantly the Value-At-Risk, or other risk measures (e.g. TVaR)
  - It increases the ability of the company to survive extreme events
- In other words it reduces the need in Solvency Capital
- Hence the name of diversification benefit

And dependencies

- When risks are independent (as r.v.), the diversification benefit goes to 100% as the number of risks go to infinity
- However insurance risks are clearly not independent from each other
- In this case, even infinitely large portfolios cannot diversify fully
- Dependencies between risk = less diversification benefit

Introduction (2)

Importance of an adequate modelling

- Although dependencies lower the diversification benefit, diversification benefit still is the major source of capital release for large portfolios.
- E.g. SCOR Non-Life portfolio profit from diversification up to 75%
- Therefore, a poor estimation and/or a poor modelling of dependencies may lead to a disastrous under/overestimation of Solvency Capital Requirement.

Modelling dependencies

- Is an highly non trivial task
- Individual risks have in general quite well-known distribution function from historical data (except rare events, e.g. rare earthquakes)
- But the joint distribution of these risks is much less known, and requires some modelling, assumptions, and calibration.
- One way of taking into account dependencies is by tying together these individual risks into a joint distribution, via a function called a copula

Diversification benefit in Gaussian Aggregation Trees

Modelling dependencies with copula-based aggregation trees (1)

The linear correlation approach

- One could simply consider and calibrate all the linear correlation coefficients between all the individual risks to get a picture of the behavior of the full portfolio
- **Major drawback:** for large portfolios the method is impossible to apply in practice (1000 risks => half a million of unknown parameters)
- **Moreover,** linear correlation coefficients do not carry all the information about the dependency structure

The Copula approach

- **Sklar's Theorem** 1959, in short:

For any continuous joint distribution function F between r.v. Xi with marginals d.f. F_i, there exists one and only function C such that

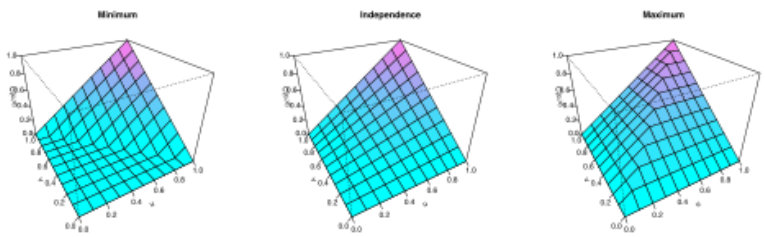
$F(x_1,\ldots x_n)=C(F_1(x_1),\ldots, F_n(X_n))$

And the converse holds iff some conditions on C are met

- **Intuitively, the copula C grasps all the information on the dependency between the r.v.**

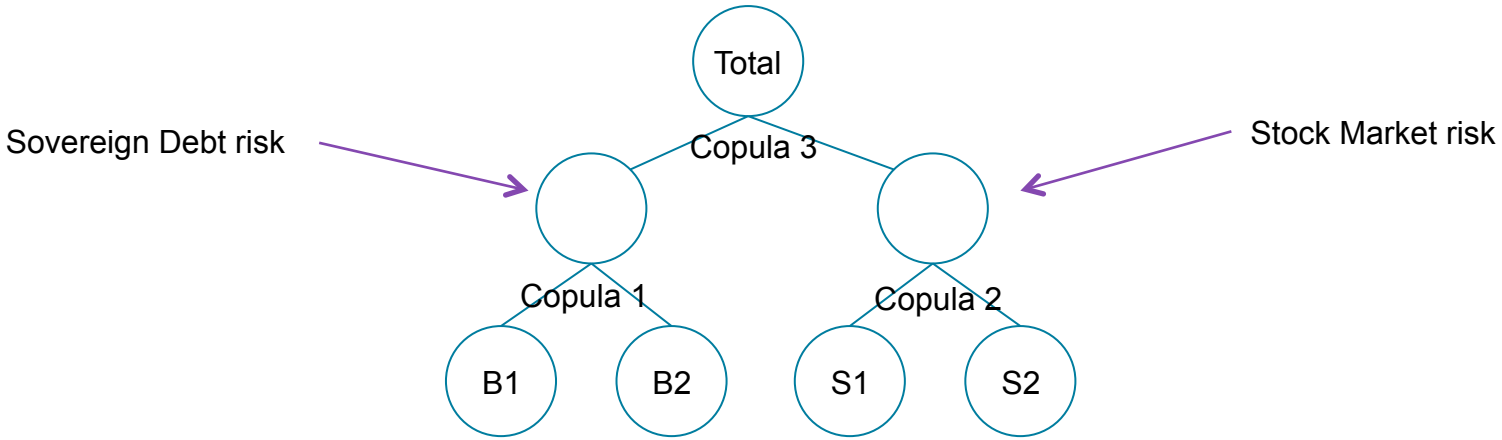
Which copula?

- **There exists infinitely many copulas**; in particular several one (or more) parameter families of copulas: Elliptical Copulas (Gaussian, Student's t), Archimedean copula (Clayton, Gumbel…), etc.
- They typically capture better the tail-dependencies (dependencies between extreme events) than linear coefficients
- While setting the dependencies between risks, **there are modelling choices!**



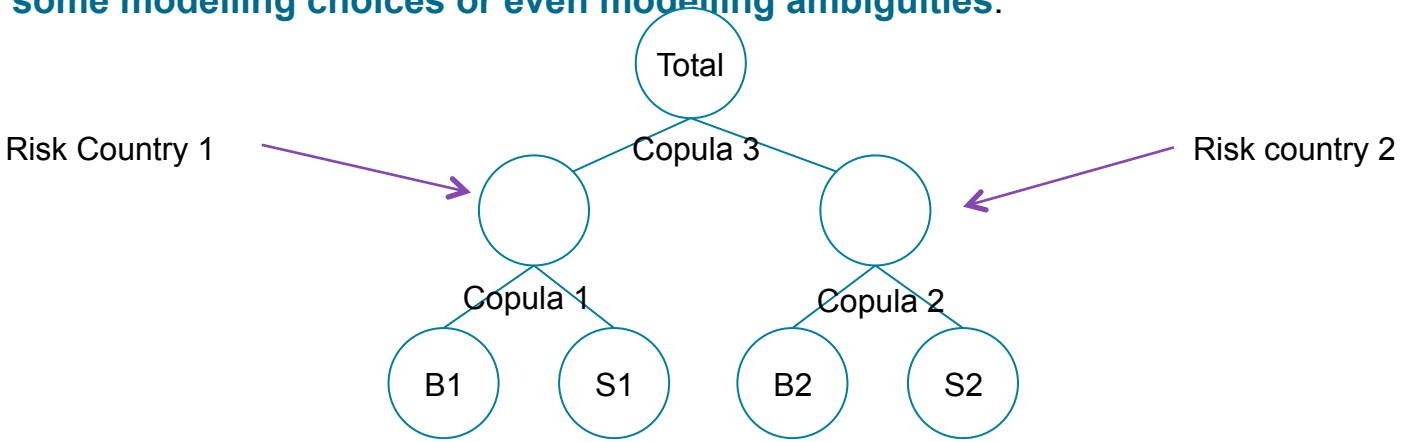
Modelling dependencies with copula-based aggregation trees (2)

- Again, it would make little sense, and would be very difficult in parctice to find a copula joining together many different risks
- Instead, a common practice which actually fits better the business is to aggregate risks hierarchically.
- Risks of “common nature” are linked together via a one parameter copula (simplification)
- This defines subset of risks aggregated together, and the procedure repeats to upper levels.
- What we are interested in is the distribution function (the VaR or TVAR) of the full portfolio $Z=\Sigma$ Xi which will depend on the aggregation mechanism (copula chosen, parameters, order of aggregation)
- An exemple: Two countries 1 and 2, four assets: Stocks in country 1 and 2, Governments Bonds in country 1 and 2



Advantages and drawbacks

- **Great reduction** of the number of parameters. In the example, 3 parameters instead of 10 linear correlation coefficients. For larger (and more realistic trees), approx. hundred of parameters vs. 10^6!
- **Still some modelling choices or even modelling ambiguities**:



- … But the structure itself enables one to model the aggregation in such a way that it reflects the business
- While building such trees, it is critical to stay as close as possible to the reality: in particular in regards of the structure of the tree itself, and the dependencies (copulas) and their calibration.
- At SCOR, there are dedicated teams working on the issue.

Diversification benefit in Gaussian Aggregation Trees

Formal Definition

- The definition of diversification benefit requires a risk measure. We will use the xTVaR: Tail-Value-at-Risk or Expected Shortfall, minus the mean, at some threshold α
- (The TVaR or xTVaR have many advantages over the widely used VaR)
- Then we define:
  - S_Z= xTVaR(Z) $S_1 = \Sigma$ xTVaR(X_i) S_0= xTVaR(ΣX_i)$|_{no\ dep}$

- S_Z is the expected shortfall (minus the mean) of the full portfolio: so this is the actual sum at risk.
- S_1 instead is the sum of the expected shortfall of all individual risks. It is the sum at risk when diversification is zero, i.e. when all risks are fully dependent between each other
- S_0 is another fictive sum at risk, where dependencies are set to zero, i.e. when diversification is maximal.
- Diversification benefit is usually defined as: $DB = 1 - S_Z / S_1$
- Noting that we always have $S_0 \leq S_Z \leq S_1$
  We also define the *diversification factor* η as $\eta = (S_Z - S_0) / (S_1 - S_0)$
- η ranges from 0 to 1, and measures how close is the total portfolio to the full dependency case (no diversification and η =1), or to the zero dependency case (maximal diversification and η=0)

Diversification benefit in Gaussian Aggregation Trees

Motivation for studying the Gaussian Aggregation Tree

Motivation:

- Our goal was to understand the behavior of the diversification benefit as a function of the shape of the tree.
- We also wanted to understand why the DB can be high even if strong dependencies are applied at each steps of the aggregation.
- Intuitively, this is because far away nodes in the tree are joined together via several times the application of the copulas, therefore “diluting” the dependencies.
- The study of the Gaussian Tree is then motivated because it can be solved exactly, and we get explicit analytical formulas for DB and η as a function of the tree parameters (width, depth, and number of leaves)

Setup

- To get rid of the effect of the order of aggregation (and also for simplicity), we consider that all individual risks are Gaussian with zero mean and the same variance σ
- We only study *regular trees*, where a node always has *k* children. The root of the tree is at level 0, the leaves are at level *m*
- Therefore there are $N=k^m$ leaves in such a tree.
- Risks are aggregated by groups of *k* nodes to the immediate parent in the tree, via the equicorrelation k-dim Gaussian copula of dependency parameter ρ.
- Meaning that we assume these groups of k nodes to be jointly normal with the covariance matrix

$$\begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}$$

Brief derivation of the main result

- First step of the aggregation: we sum together N/k groups k gaussians which are jointly normal (because of the Gaussian copula) .
- A standard theorem on multivariate normal then states that the sum of these k r.v. is also a Gaussian, and its variance is given as a function of *k*, ρ, and σ

$$\sigma\downarrow upperlevel = \sigma\sqrt{k+(k\uparrow 2-k)\rho}$$

- At the next level, we then have again only gaussians, and the procedure therefore iterates straightforwardily
- The result is that the full portfolio is also a Gaussian, with variance given by

$$\sigma\downarrow Z = \sigma\,(k+(k\uparrow 2-k)\rho)\uparrow m/2$$

- We then use some standard results on the TVaR of gaussians, and derive the different sum at risk S_0,S_1, and S_Z. Then:

$$DB(k,m,\rho) = 1-\left(\frac{1}{k}+\left(1-\frac{1}{k}\right)\rho\right)^{m/2}$$

$$\eta(k,m,\rho) = \frac{\left(k+(k^2-k)\rho\right)^{m/2}-k^{m/2}}{k^m-k^{m/2}}$$

Behavior of the diversification factor (1)

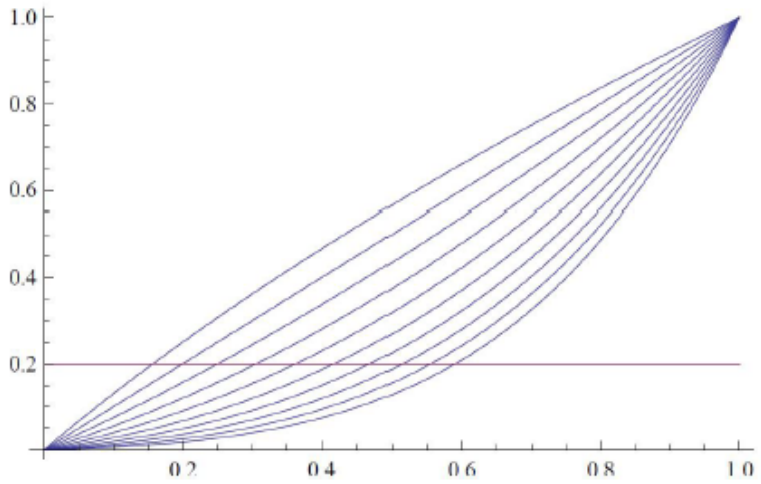


Figure 3.1: The behaviour of η (y-axis) with ρ (x-axis) in a m-level Gaussian tree. Here k is set to 3 and m ranges from 1 (top curve) to 10 (bottom curve). The $m = 1$ roughly behaves as $\sqrt{\rho}$, as explained in Chapter 2. However, the curves become convex for $m \geq 3$, hence giving a small η even if ρ is large. The straight line is the observed value for $\eta \approx 0.2$. We see that a Gaussian tree with $m = 6$, $k = 3$, $N = 729$ can fit the value of η for a dependency parameter given by $\rho \approx 0.4$ (see the 6^{th} curve from the top)

Increasing the depth of the tree makes more concave the curve for η, meaning that the diversification can be large although ρ is high

The depth becomes large enough to lower significantely the effective dependencies!

Behavior of the diversification benefit (2)

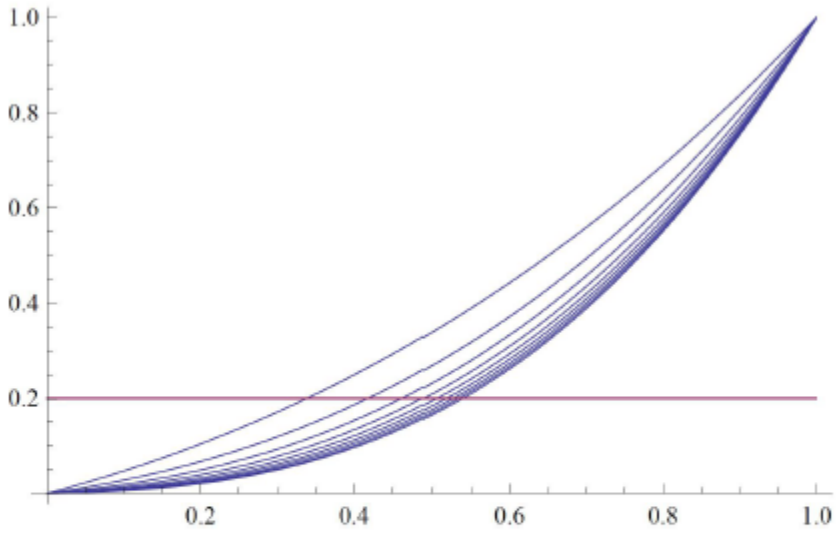


Figure 3.2: The behaviour of η (y-axis) with ρ (x-axis) in a m-level Gaussian tree, with $m = 6$ and k ranging from 2 (top curve) to 10 (bottom curve). Given a depth of the tree, increasing the number of children k increases the total number of leaves and the η curve approaches a limiting curve given by $\rho^{m/2}$.

"Tight" Trees diversify better than "fat" trees (1)

- The discussion on the depth of the tree is not relevant enough, since changing *m* also changes the total number of individual risks.
- But one might work at fixed N.

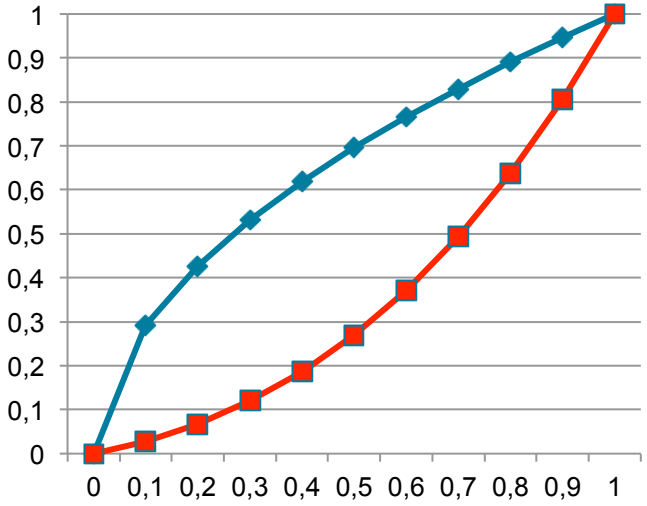
E.g:

- $k = 2$, $m = 10$, $N = 1024$. Then we $DB \approx 83\%$ and $\eta \approx 0.14$
- $k = 4$, $m = 5$, $N = 1024$. In this case, $DB \approx 77\%$ and $\eta \approx 0.2$

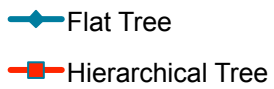
- The DB is higher for the tighter tree (k=4, m=5).
- This can be extended in full generality to any two trees k,m and k',m' such that N=N'. Hence the proof of the claim.
- The "fattest" tree is the flat tree (m=1, N=k), whereas the tightest is defined by k=2

“Tight” Trees diversify better than “fat” trees (2)

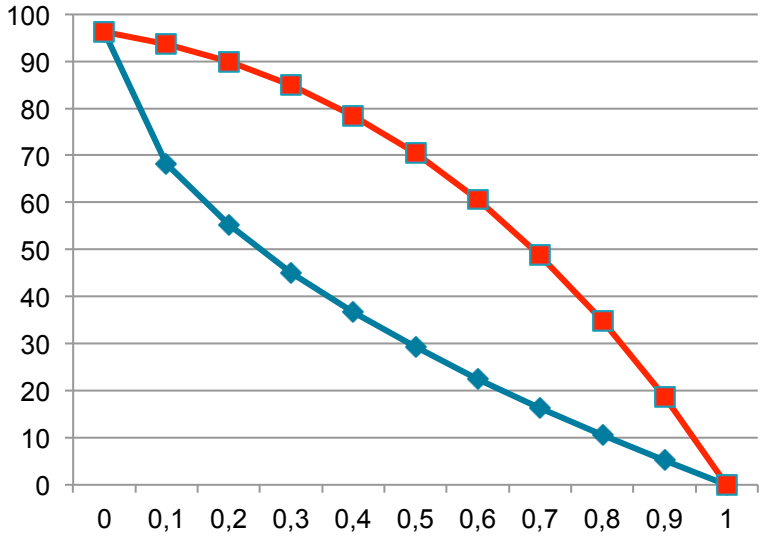
- Here we compare two Gaussian trees with N=729.
  - Flat Tree has k=729,
  - Hierarchical Tree has k=3, m=6



Diversification factor



and



Diversification benefit

Diversification benefit in Gaussian Aggregation Trees

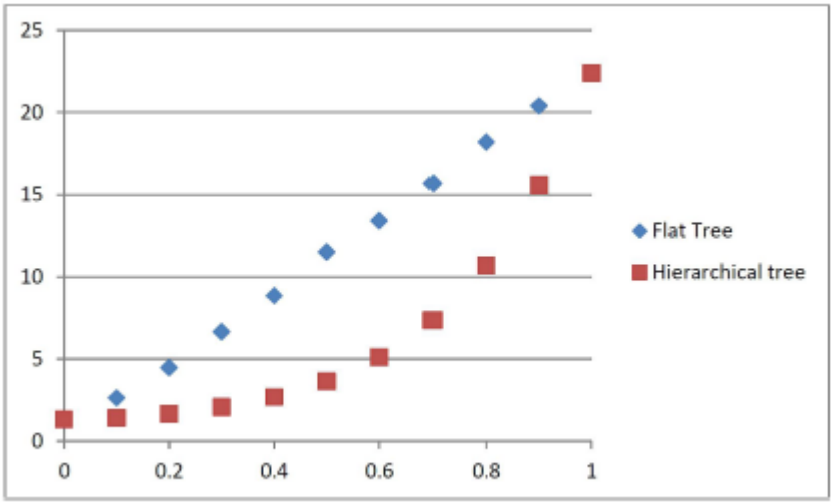
LogNormal Trees

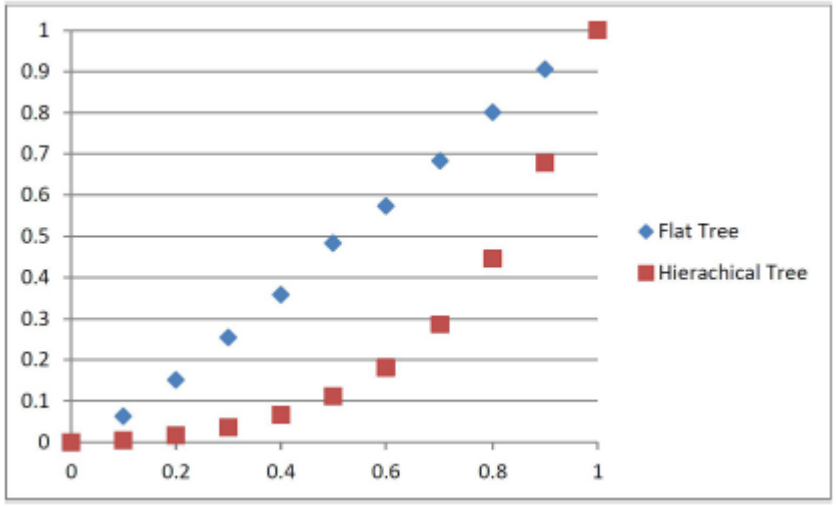
- We test our previous result on more relevant (fat-tail) distributions
  - The individual risks are assumed to be given by the same LogNormal, parameterized by its mean and variance
  - The aggregation is done via Gaussian copula of parameter ρ, or a Clayton Copula

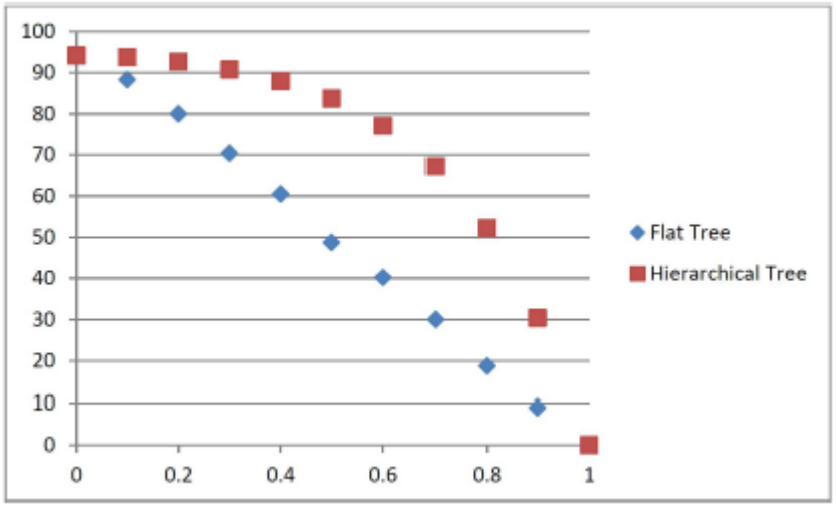
- The parameterization has been chosen in order to reflect the actual Non-Life portfolio of SCOR:
  - Mean of LogNormal is approx 670'000
  - STDEV is around 8.1 million
  - There are 3^6=729 individual risks
  - Then S_0 is 1.3 bn and S_1 is 22.4bn
  - The S_Z coming from the evaluation of the SCOR's tree is around 5bn, meaning
    - DB(observed) approx 78%
    - η (observed) approx 0.17

- We now run MC simulations on our toy-model tree using IGLOO software both for flat and hierarchical tree (k=3,m=6)

Diversification in LogNormal Trees with Gaussian copula





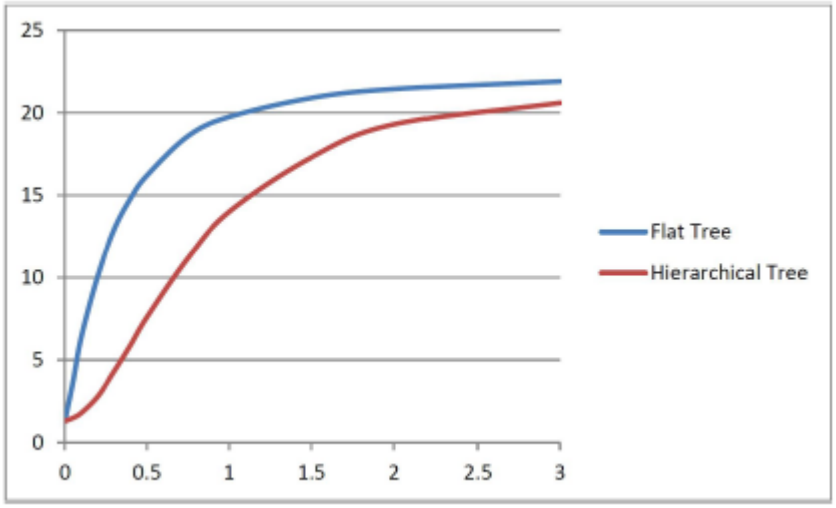


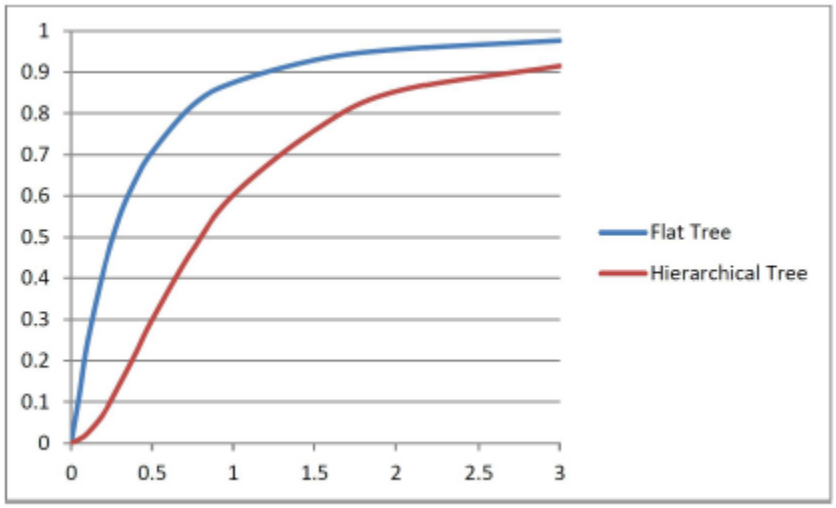
t.l: Sum At Risk; t.r. eta factor, b.l. diversification benefit

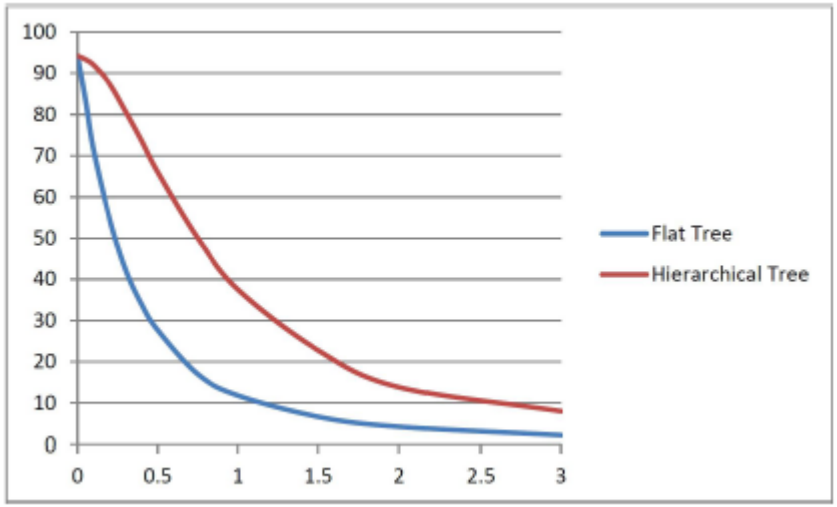
As in the Gaussian case, hierarchical tree makes the curve for η concave, or the curve for DB more convex. In other words the DB stay very close to the fully independent case until quite high values for ρ.

Also, the behavior is clearly non-linear, meaning that the sensitivity of DB to ρ is high around realistic values (ρ approx 0.6). This is a drawback of the hierarchical tree. E.g. from 0.7 to 0.8, DB is down by 15%!

Diversification in LogNormal Trees with Clayton copula







t.l: Sum At Risk; t.r. eta factor, b.l. diversification benefit

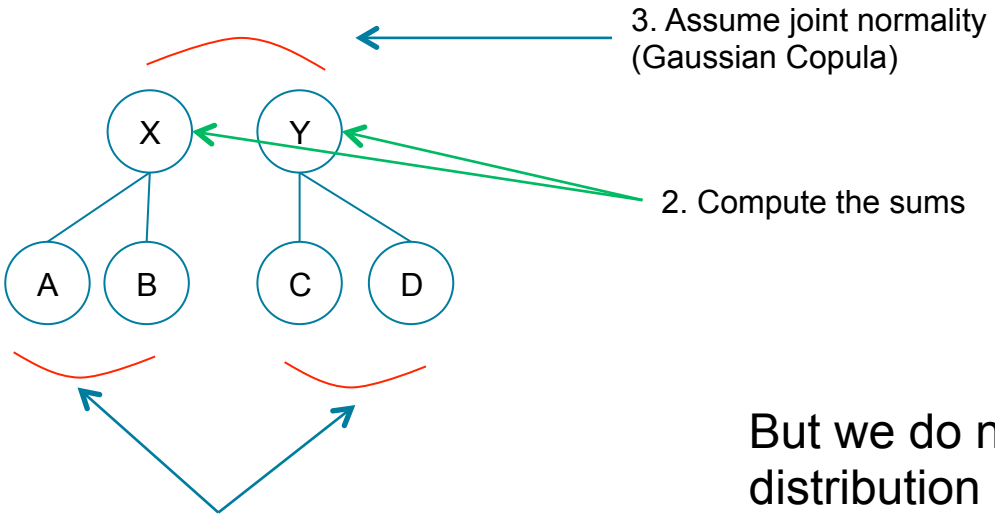
Now the hierarchical aggregation fails to turn the curve into concave curve for η or convex curve for DB, but still the HT diversifies much more than the flat tree. The toy model fits observed values with a Clayton parameter of approx 0.5

Again the sensitivity of DB to Clayton parameter is very high around realistic values (0.5).

Diversification benefit in Gaussian Aggregation Trees

Tree of aggregation: mathematical background (1)

- The way we presented the mechanism of hierarchical aggregation was enough in order to derive the variance of the full portfolio, but *is not enough to specify fully* « the tree », i.e. to specify the joint distribution of all the leaves!
- The aggregation mechanism we used was:
  - Assume the leaves are jointly normal by groups of k nodes.
  - Assume their sum is again jointly normal, and so on.



But we do not know the joint distribution of these four leaves!

Tree of aggregation: mathematical background (2)

- Given individual risks Xi, we are interested in the total portfolio Z=Σ Xi
- If the joint distribution is known (equivalently the joint characteristic function), then the sum might be computed as:

$$\Phi_{\mathbf{X}}(t_1, t_2, \ldots, t_n) = \mathbb{E}[\exp(i\mathbf{t}.\mathbf{X})]$$
$$\Phi_Z(t) \equiv \mathbb{E}\left[\exp\left(itZ\right)\right] = \mathbb{E}\left[exp\left(it\left(\sum X_i\right)\right)\right]$$
$$\Phi_Z(t) = \Phi_{\mathbf{X}}(t, t, \ldots, t)$$

- But the joint distribution is unknown in general, and one may only have limited amount of information. For instance

$$\mathbf{Y_1} = (X_1, \ldots, X_j)$$
$$\mathbf{Y_2} = (X_{j+1}, \ldots, X_n)$$
$$\Phi_Z(t) = \mathbb{E}\left[\exp\left(it\left(\sum X_i\right)\right)\right] = \mathbb{E}\left[\exp\left((i\left(tY_1 + tY_2\right)\right] \equiv \Phi_{Y_1,Y_2}(t,t)$$

- Then one is able to compute Z even if the joint distribution of the Xi's is unknown, but both Y1, Y2, and their joint distribution are.
- One could then further split Y1, etc.
- This top down decomposition of the sum exactly corresponds to the bottom-up aggregation of risks we have considered, and is indeed well represented by a tree.
- The main point about tree aggregation is thus that the joint distribution of all the leafs is not needed to be able to compute the sum!

How to specify fully the tree?

- The graphical representation suggested, as in Bayesian networks/directed acyclic graphs that conditional independence statements could be added to specify uniquely the tree.

- We considered *conditional independence statements,* of the form: A $\perp\!\!\!\perp$ C | X, and more generally

$$\forall Y \in T, \forall X \leq Y, \forall S \not\leq Y, (X \perp\!\!\!\perp S)|Y$$

- Then the aggregation mechanism leads to a unique solution.

Theorem 2: There exist one and only one regular (k, m) Gaussian tree $T_{k,m}$ constructed the way specified above, and that is at the same time conditional independent. In this case, the k^m leaves of the tree are jointly normal with a unique covariance matrix $C^{(m)}$ given by the following recurrence relations:

$$C^{(0)} = (\sigma_Z^2)$$

$$C^{(1)} = \sigma_{(1)}^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}$$

And, $\forall m \geq 2$:

$$C^{(m)} = \sigma_{(m)}^2 \begin{pmatrix} \frac{C^{(m-1)}}{\sigma_{(m-1)}^2} & B'_{(m-1)} & \cdots & B'_{(m-1)} \\ B_{(m-1)} & \frac{C^{(m-1)}}{\sigma_{(m-1)}^2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & B'_{(m-1)} \\ B_{(m-1)} & \cdots & B_{(m-1)} & \frac{C^{(m-1)}}{\sigma_{(m-1)}^2} \end{pmatrix}$$

How to specify fully the tree?

where $\mathcal{C}^{(m)}$ is here written as a $k \times k$ block matrix, with blocks of size k^{m-1}, and where

- The recurrence relations between the variances is:

$$\sigma_{(p-1)} = \sigma_{(p)} \sqrt{k + (k^2 - k)\rho}$$

In particular, one has

$$\sigma_Z = \sigma_{(m)} \left(k + (k^2 - k)\rho\right)^{m/2}$$

- The matrix $B_{(m-1)} = \beta_{m-1} J_{k^{m-1}}$ with

$$\beta_{m-1} = \rho \left(\frac{1}{k} + \left(1 - \frac{1}{k}\right)\rho\right)^{m-1}$$

- Proof: lengthy

Corollary: effective couplings

- This unique CI regular Gaussian tree enables us to define effective coupling in the tree. We call effective coupling between two individual risks Xi and Xj (leaves) ρ_{ij}=Cov(Xi,Xj)

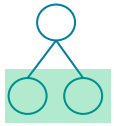
- It is easily shown to depend only on the level *p* at which one finds the first node connecting these two, and we derive (from the full covariance matrix), the result:

$$\rho_{\text{eff}}^{(m,p)} = \rho \left(\frac{1}{k} + \left(1 - \frac{1}{k} \right) \rho \right)^{m-p-1}$$

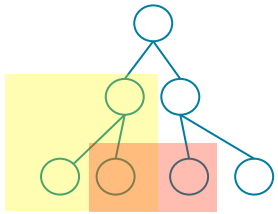
- This shows that the correlation between individual risks decreases with their “distance” in the tree, and explains why hierarchical trees have the tendency to lower significantly the overall interdependency.

Effective couplings: An example with k=2 and ρ=0.2

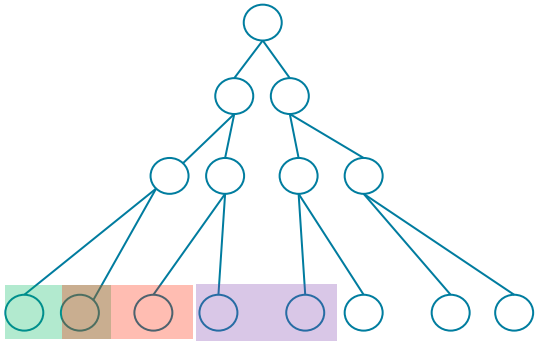
$$\mathcal{C} = \sigma^2 \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$$



$$\mathcal{C} = \sigma^2 \begin{pmatrix} 1 & 0.2 & 0.12 & 0.12 \\ 0.2 & 1 & 0.12 & 0.12 \\ 0.12 & 0.12 & 1 & 0.2 \\ 0.12 & 0.12 & 0.2 & 1 \end{pmatrix}$$



$$\mathcal{C} = \sigma^2 \begin{pmatrix} 1 & 0.2 & 0.12 & 0.12 & 0.07 & 0.07 & 0.07 & 0.07 \\ 0.2 & 1 & 0.12 & 0.12 & 0.07 & 0.07 & 0.07 & 0.07 \\ 0.12 & 0.12 & 1 & 0.2 & 0.07 & 0.07 & 0.07 & 0.07 \\ 0.12 & 0.12 & 0.2 & 1 & 0.07 & 0.07 & 0.07 & 0.07 \\ 0.07 & 0.07 & 0.07 & 0.07 & 1 & 0.2 & 0.12 & 0.12 \\ 0.07 & 0.07 & 0.07 & 0.07 & 0.2 & 1 & 0.12 & 0.12 \\ 0.07 & 0.07 & 0.07 & 0.07 & 0.12 & 0.12 & 1 & 0.2 \\ 0.07 & 0.07 & 0.07 & 0.07 & 0.12 & 0.12 & 0.2 & 1 \end{pmatrix}$$



Further work

- What is the minimal set of CI statements that are needed to specify completely the tree (Gaussian or not)?
- What is the relation to yet another algorithm of aggregation : “join *k* nodes together, get *k* such multivariate, join them again together, …” and take the sum only at the end. This can be described by copulas of copulas of copulas, etc (nested copulas).
- Does this last algorithm correspond better to what softwares are doing? Also, does this automatically imply the CI statements?
- These are open questions…

Conclusion

- We studied a toy model of hierarchical aggregation tree of risks whose individual marginals are known. We assumed both Gaussian marginals and Gaussian copula, and showed that the total portfolio remains a Gaussian whose variance is easily derived.
- This enables to compute exactly the DB and study the influence of the shape of the tree.
- Hierarchical trees have a strong tendency to erase the dependencies, so that the DB stays actually high and almost flat for a wide range for the dependency parameter!
- One should therefore be cautious about the amount of DB, and justify in any case the modelling of the risk aggregation tree, in particular its shape.
- This is indeed relevant as it does also apply to more realistic trees such as trees with Lognormal risks and Clayton copulas.
- Regarding the maths side, the “why” of a tree of aggregation is understood as a decomposition of the sum that fits with the available information; a natural consequence is that it does not specify uniquely the joint distribution
- Further work is needed on the minimal set of extra requirements (like CI statements), and if these ones might naturally come from more involved algorithms of aggregation (e.g. nested copulas)