Modelling dependencies between internal and external models using vine copulas

Tim Harrison *

2025-08-20

Abstract

Dependencies are a major theme in capital modelling, and these are often some of the most material and subjective judgments in an internal model. Most internal models rely to some extent on external models, and dependencies between internal and external models are increasingly important, given the growing tendency for risks in one area of the risk profile to impact outcomes in other risk areas; the impact of economic conditions on underwriting performance, and the importance of non-physical climate-change risks are just some examples.

This paper provides a novel method, which goes part way to addressing a limitation in apparent standard practice, by enabling the specification of more than one dependence relationship between sets of pre-simulated data. It therefore sets actuaries up to be able to deal with the ever-more correlated world around us.

1 Introduction

Insurance firms operating under regimes such as Solvency II have the option to calculate their Solvency Capital Requirements (SCR) using an approved internal model, which in turn often relies on Monte Carlo simulation techniques to produce a distribution of future outcomes. These internal models often make use of outputs from so-called external models such as Economic Scenario Generators (ESGs) or Natural Catastrophe models. External models are used to supplement the internal model in areas where the firm's own internal data, expertise, or resources are not sufficient to design, calibrate, and maintain an appropriate model.

The output from external models arrives 'pre-simulated' from the point of view of the internal model, in that it is usually a table of fixed numbers with a column for each variable and a simulation trial in each row. The joint distribution implicit in the external model output must be preserved—in other words, the columns in the table cannot be reordered independently of one another, and the consistency of each row in the table must be maintained.

This presents a challenge, as we often wish to specify and model various dependence relationships between variables in the external models and the internal model. This is becoming increasingly important given the interconnected nature of the modern world, the links between financial markets and physical phenomena, and last but not least the growing impact of climate change uncertainty. However, the only way that we can induce dependencies with a set of pre-simulated figures is to change their order relative to the variables in the internal model.

*Liberty Specialty Markets,tim.harrison@libertyglobalgroup.com

To model a dependency between a set of external model output and the internal model, apparent standard practice is to specify a single bivariate copula between one variable in the pre-simulated data and one in the internal model. This has various limitations, not least of which is that this can result in unintended and difficult-to-control dependencies between other pairs of variables, but no parameters to control these relationships.

This article therefore introduces a novel method which goes some way to resolving this problem, allowing actuaries to specify and achieve additional dependencies between pairs of variables in sets of pre-simulated data and the internal model.

2 Apparent industry standard approach

In this section, we establish what I refer to as the industry standard approach and demonstrate a major limitation inherent in its use.

Let $\mathbf{X}_A$ be a set of pre-simulated variables from an external model, with k observations of m variables $\{x_1, \ldots, x_m\}$ and correlation matrix R_A. We wish to preserve the joint distribution of $\mathbf{X}_A$ exactly, and we therefore will not re-order the sample relative to itself. This may be because the sample is drawn from an external model such as an ESG or a Natural Catastrophe model.

Let $\mathbf{X}_B$ be a set of pre-simulated data from an internal model with k observations of $(n-m)$ variables $\{x_{m+1}, \ldots, x_n\}$ and correlation matrix R_B. As above, $\mathbf{X}_B$ has a joint distribution which we want to preserve.

Assume that the samples for $\mathbf{X}_A$ and $\mathbf{X}_B$ are generated independently. In the event that we make no further assumptions, the joint density can then be written as the product of the densities:

$$f_{AB}(x_1, \ldots, x_n) = f_A(x_1, \ldots, x_m)\, f_B(x_{m+1}, \ldots, x_n)$$

If we wish to induce dependence between variables in $\mathbf{X}_A$ and $\mathbf{X}_B$, industry standard practice is to specify a single bivariate copula between one pair of variables $\{X_i, X_j\}, i \in A, j \in B$. A copula C is a distribution on the unit square with standard uniform marginals, and via Sklar's theorem [1] a joint distribution can be expressed in terms of a copula distribution function C as follows:

$$F_{1\ldots n}(x_1, ..., x_n) = C_{1\ldots n}[F_1(x_1), ..., F_n(x_n)]$$

with joint density:

$$f_{1\ldots n}(x_1, ..., x_n) = c_{1\ldots n}[F_1(x_1), ..., F_n(x_n)] \cdot \prod_{i=1}^{n} f_i(x_i) \tag{1}$$

If we specify a bivariate copula with distribution function C_{ij} and density c_{ij} between some $X_i, i \in A$ and some $X_j, j \in B$ then the joint density becomes:

$$f_{AB}(x_1, \ldots, x_n) = f_A(x_1, \ldots, x_i, \ldots, x_m)\, f_B(x_{m+1}, \ldots, x_j, \ldots, x_n)\, c_{ij}[F_i(x_i), F_j(x_j)]$$

All other dependencies between variables in $\mathbf{X}_A$ and $\mathbf{X}_B$ are not specified, and therefore an implicit assumption of conditional independence is being made.

For a set of simulation output, this dependence could be achieved in practice by re-ordering the two samples relative to one another, e.g. by the following algorithm:

1. Given samples for $\mathbf{X}_A$ and $\mathbf{X}_B$ with k observations, sample k observations $\{u, v\}$ from C_{ij}.
2. Re-order the sample of $\mathbf{X}_A$ such that the ranks of x_i and u are equal.
3. Re-order the sample $\mathbf{X}_B$ such that ranks of x_j and v are equal.
4. Hence x_i and x_j have bivariate copula C_{ij}.

The correlation matrix for all variables R_{AB} can be written in block form as:

$$R_{AB} = \begin{pmatrix} R_A & Y \\ Y^T & R_B \end{pmatrix}$$

Where Y is the incompletely-specified matrix of inter-model correlations and only ρ_{ij} is specified:

$$Y = \begin{pmatrix} \rho_{1,m+1} & \rho_{1,m+2} & \cdots & \rho_{1,j} & \cdots & \rho_{1,n} \\ \rho_{2,m+1} & \rho_{2,m+2} & \cdots & \rho_{2,j} & \cdots & \rho_{2,n} \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ \rho_{i,m+1} & \rho_{i,m+2} & \cdots & \rho_{i,j} & \cdots & \rho_{i,n} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ \rho_{m,m+1} & \rho_{m,m+2} & \cdots & \rho_{m,j} & \cdots & \rho_{m,n} \end{pmatrix}$$

To find the unspecified values, by perturbing the rows in the matrix we can rewrite the correlation matrix in block form, where R_{ii} is a known symmetric matrix with ones on the diagonal, and where the values for $\mathbf{b}$ and $\mathbf{g}$ are known, as:

$$R_{AB} = \begin{pmatrix} 1 & \mathbf{b} & \mathbf{c} & d = \rho_{ij} \\ \mathbf{b}^T & R_{22} & E & \mathbf{f} \\ \mathbf{c}^T & E^T & R_{33} & \mathbf{g} \\ d & \mathbf{f}^T & \mathbf{g}^T & 1 \end{pmatrix}$$

and using the results in [2] find the unspecified values in R_{AB} as:

$$\begin{aligned} \mathbf{c} &= \rho_{ij}\, \mathbf{g}^T \\ \mathbf{f} &= \mathbf{b}^T \rho_{ij} \\ E &= \mathbf{f}\, \mathbf{g}^T \end{aligned} \tag{2}$$

It is clear that all unspecified terms involve ρ_{ij}. The limitation of the industry standard approach, therefore, is that while we may wish to specify additional pairwise dependence relationships, the model has no parameters to control them. In fact, in many commonly used capital modelling software platforms, the software itself will enforce an assumption of conditional independence and restrict the specification of any further dependence relationships between other pairs of variables in $\mathbf{X}_A$ and $\mathbf{X}_B$.

We therefore seek an alternative approach which allows for the specification of a more flexible set of dependence assumptions between sets of pre-simulated data, achieved only by re-ordering the samples relative to one another.

3 Vine copulas

The alternative approach proposed later in this paper, to extend the industry standard approach discussed above, relies on the idea of vine copulas. As vine copulas are not yet widely used in actuarial practice, this section introduces the idea of vines and some other related concepts. First, we introduce pair copula decomposition of a joint density.

3.1 Pair copula decomposition of a joint density

A joint density for n variables can be factorised into the product of conditional densities as follows:

$$f_{1\ldots n}(x_1, \ldots, x_n) = f_n(x_n)\, f_{n-1}(x_{n-1}|x_n) \cdots f_1(x_1|x_2, \ldots, x_n)$$

Recalling (1), in the bivariate case we have:

$$f_{12}(x_1, x_2) = f_2(x_2)\, f_{1|2}(x_1|x_2) = c_{12}[F_1(x_1), F_2(x_2)]\, f_1(x_1)\, f_2(x_2)$$

and hence:

$$f_{1|2}(x_1|x_2) = c_{12}[F_1(x_1), F_2(x_2)]\, f_1(x_1)$$

In the case of 3 random variables, by iterating the same approach, we can write:

$$\begin{aligned} f_{123}(x_1, x_2, x_3) &= f_3(x_3) \times f_{2|3}(x_2|x_3) \times f_{1|23}(x_1|x_2, x_3) \\ &= f_3(x_3) \times c_{23}[F_2(x_2), F_3(x_3)]\; f_2(x_2) \\ &\quad \times\; c_{13|2}[F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3)]\; c_{12}[F_1(x_1), F_2(x_2)]\; f_1(x_1) \end{aligned} \tag{3}$$

Or in other words, a joint density can be decomposed into a product of marginal densities and pair copula densities, of which some are conditioned. In general, there are many pair-copula decompositions of any joint distribution, and so we ideally need a way to structure our approach. For this we rely on the idea of a vine copula.

3.2 General definition for regular vines

As we saw in (3), a pair copula decomposition results in a set of conditioned pair copulas. Regular vines provide a systematic way for us to find possible decompositions for a given joint distribution.

A vine is just a graphical tool for identifying and labeling bivariate constraints in a pair-copula decomposition. Vines are sets of trees, and each tree represents a set of pairwise dependence relationships.

Via Bedford & Cooke [3] we have the following definition for a regular vine:

- A vine $\mathcal{V}$ on n variables is a set of trees $\{T_1, ..., T_{n-1}\}$.
- A tree T_i is an undirected, acyclic graph with nodes N_i and edges E_i.
- T_1 is a tree with nodes $N_1 = \{1, ..., n\}$ and edges E_1.
- For $i = 2, ..., n-1$, T_i is a tree with nodes $N_i = E_{i-1}$, i.e. the nodes in tree T_{i-1} are the edges in tree T_i.
- For a vine to be *regular* then for $i = 2, ..., n-1$ if $\{a, b\} \in E_i$ we need that $|a \cap b| = 1$, i.e. for nodes to be connected in T_i they must share a node in T_{i-1}.

To find the conditional constraint for edge $e = \{a, b\} \in E_i$:

- The constraint set is the complete union U_e of e, i.e. the subset of $\{1, ..., n\}$ reachable from e and found as the union of the constraint sets of the nodes: $U_e = U_a \cup U_b$.
- The conditioning set is the intersection of the constraint sets of the nodes: $D_e = U_a \cap U_b$.
- The conditioned set (always a doubleton) is found as: $\{L_{e,a}, L_{e,b}\} = \{U_a \setminus D_e, U_b \setminus D_e\}$.

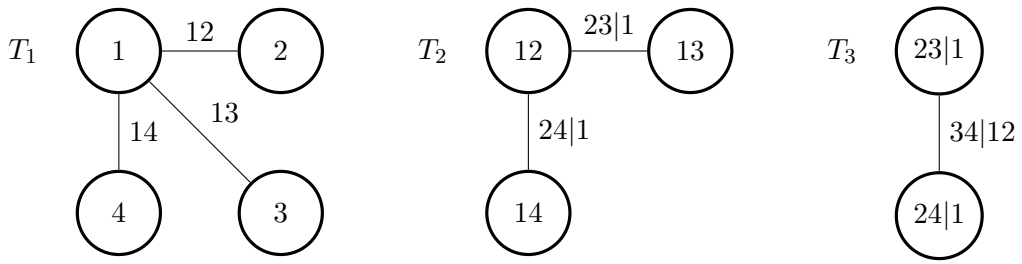

Figure 1: A regular vine with four variables.

The joint density for n-dimensions then takes the following general form, as the product of marginal densities f and pair copula densities c appropriately indexed:

$$f_{1\ldots n}(x_1, ..., x_n) = \prod_{i=1}^{n} f_i(x_i) \prod_{i=1}^{n-1} \prod_{e \in E_i} c_{L_{e,a}, L_{e,b}|D_e}[F(x_{L_{e,a}}|x_{D_e}), F(x_{L_{e,b}}|x_{D_e})]$$

3.3 Example of a vine

The formal definition may seem intimidating at first glance, so let us briefly consider an example with 4 variables, as shown in Figure 1. The key point is to understand how a vine helps us to understand what conditional constraints in higher-level trees are implied by our choices about the graph in lower-level trees.

Looking at the graph of Figure 1 and applying the definition above:

$$\begin{array}{lll}
\mathcal{V} = \{T_1, T_2, T_3\} & & \\
T_1 = \{N_1; E_1\} & N_1 = \{1,2,3,4\} & E_1 = \{(1,2),(1,3),(1,4)\} \\
T_2 = \{N_2; E_2\} & N_2 = E_1 & E_2 = \{\big((1,2),(1,3)\big), \big((1,2),(1,4)\big)\} \\
T_3 = \{N_3; E_3\} & N_3 = E_2 & E_3 = \{\big(((1,2),(1,3)),((1,2),(1,4))\big)\}
\end{array}$$

Tree $\mathbf{T_1}$
For the edges in E_1 we have the following:

$$\begin{aligned}
U_{(1,2)} &= \{1,2\} \\
U_{(1,3)} &= \{1,3\} \\
U_{(1,4)} &= \{1,4\}
\end{aligned}$$

The edges in E_1 have empty conditioning sets and can be said to correspond to the pair copulas between variable X_1 and each of X_2, X_3, and X_4:

$$\begin{aligned}
(1,2) &\longleftrightarrow C_{12}[F_1(x_1), F_2(x_2)] \\
(1,3) &\longleftrightarrow C_{13}[F_1(x_1), F_3(x_3)] \\
(1,4) &\longleftrightarrow C_{14}[F_1(x_1), F_4(x_4)]
\end{aligned}$$

Tree $\mathbf{T_2}$
We must determine the constraint set, conditioning set and conditioned set for each edge $e \in E_2$.

With $e_1 = (a, b) = \big((1,2),(1,3)\big)$ we find:

$$\begin{aligned} U_{e_1} &= U_a \cup U_b = \{1,2\} \cup \{1,3\} = \{1,2,3\} \\ D_{e_1} &= U_a \cap U_b = \{1,2\} \cap \{1,3\} = \{1\} \\ \{L_{e_1,a}, L_{e_1,b}\} &= \{U_a \setminus D_{e_1}, U_b \setminus D_{e_1}\} = \{2,3\} \end{aligned}$$

Taking $e_2 = (a, c) = \big((1,2),(1,4)\big)$ we find:

$$\begin{aligned} U_{e_2} &= U_a \cup U_c = \{1,2\} \cup \{1,4\} = \{1,2,4\} \\ D_{e_2} &= U_a \cap U_c = \{1,2\} \cap \{1,4\} = \{1\} \\ \{L_{e_2,a}, L_{e_2,c}\} &= \{U_a \setminus D_{e_2}, U_c \setminus D_{e_2}\} = \{2,4\} \end{aligned}$$

Therefore we can say that the elements of E_2 correspond to the pair copulas between X_2 and each of X_3 and X_4, conditional on X_1:

$$\begin{aligned} \big((1,2),(1,3)\big) &\longleftrightarrow C_{23|1}[F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)] \\ \big((1,2),(1,4)\big) &\longleftrightarrow C_{24|1}[F_{2|1}(x_2|x_1), F_{4|1}(x_4|x_1)] \end{aligned}$$

Tree $\mathbf{T_3}$
E_3 has a single element, so taking $e_3 = (e_1, e_2) = \big(((1,2),(1,3)),((1,2),(1,4))\big)$ we find:

$$\begin{aligned} U_{e_3} &= U_{e_1} \cup U_{e_2} = \{1,2,3\} \cup \{1,2,4\} = \{1,2,3,4\} \\ D_{e_3} &= U_{e_1} \cap U_{e_2} = \{1,2,3\} \cap \{1,2,4\} = \{1,2\} \\ \{L_{e_3,e_1}, L_{e_3,e_2}\} &= \{U_{e_1} \setminus D_{e_3}, U_{e_2} \setminus D_{e_3}\} = \{3,4\} \end{aligned}$$

We can say that the element of E_3 corresponds to the pair copula between X_3 and X_4 conditional on X_1 and X_2:

$$\big(((1,2),(1,3)),((1,2),(1,4))\big) \longleftrightarrow C_{34|12}[F_{3|12}(x_3|x_1,x_2), F_{4|12}(x_4|x_1,x_2)]$$

3.4 Canonical vine / C-Vine

The general form given above for the regular vine gives rise to some commonly-used special cases. This is primarily because the number of possible vine configurations grows very quickly as the number of dimensions increases.

A vine is *canonical* if each tree has a unique node to which all other nodes are connected. The example in Figure 1 is a C-Vine. In this case with constraint set $D = \{1, ..., j-1\}$ the joint density can be written as the following product of the marginal densities and pair copula densities:

$$f_{1\ldots n}(x_1, ..., x_n) = \prod_{i=1}^{n} f_i(x_i) \prod_{j=1}^{n-1} \prod_{k=1}^{n-j} c_{j,j+k|D}[F(x_j|x_D), F(x_{j+k}|x_D)]$$

3.5 Drawable vine / D-Vine

A vine is *drawable* if in each tree, all nodes are connected to at most two other nodes. With constraint set $D = \{k+1, ..., k+j-1\}$ the corresponding joint density is then:

$$f_{1\ldots n}(x_1, ..., x_n) = \prod_{i=1}^{n} f_i(x_i) \prod_{j=1}^{n-1} \prod_{k=1}^{n-j} c_{k,k+j|D}[F(x_k|x_D), F(x_{k+j}|x_D)]$$

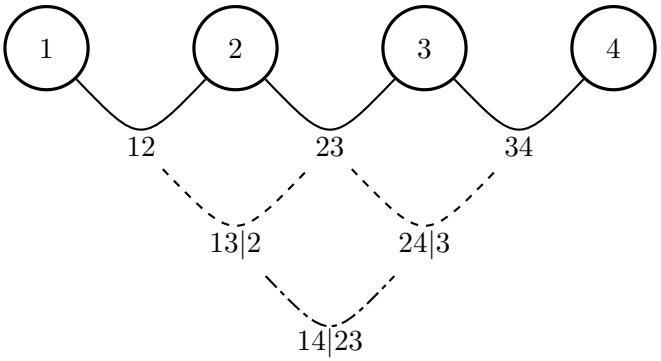

Figure 2: A D-Vine with four variables.

Figure 2 shows a 4-variable D-vine, using a slightly different representation, in which all trees are shown together, and the edges in each higher order tree clearly join *edges* in the preceding tree. This differs from the representation in Figure 1 where each tree is drawn separately. D-Vines can easily be drawn in either of these ways, which is of course where they get their name. For vines in general however, it is easier to draw each tree separately as in Figure 1.

4 Vine-based approach for correlation between internal and external models

This section presents the main result of this paper. Having briefly introduced the notion of a vine copula, we now return to the modelling problem at hand. Recall our sets of pre-simulated data $\mathbf{X}_A$ and $\mathbf{X}_B$. We wish to induce dependencies between the variables in these sets simply by re-ordering them relative to one another.

Define supplementary variable Z with density $f(z)$. By (3), we know that we can freely define one unconditional bivariate copula and $(m-1)$ conditional bivariate copulas between Z and $X_i\,, i \in A$. Let $c_{Z,i}$ be the copula density between Z and X_i, then the joint density for Z and $\mathbf{X}_A$ is:

$$f_{A,Z}(x_1,...,x_m,z) = f_Z(z) f_A(x_1,...,x_m) \prod_{i=1}^{m} c_{Z,i|1,...,i-1}[F(z|x_1,...,x_{i-1}), F(x_i|x_1,...,x_{i-1})]$$

This is equivalent to a C-Vine schema where Z is never the central node, and where we do not explicitly decompose the joint distribution of $\mathbf{X}_A$ into pair copulas, as this is the pre-simulated data whose distribution we wish to preserve.

Similarly, we can define one unconditional and $(n-m-1)$ conditional pair copulas between Z and $X_j\,, j \in B$. The joint density of Z and $\mathbf{X}_B$ can therefore be written as:

$$\begin{aligned} f_{B,Z}(x_{m+1},...,x_n,z) = f_Z(z)\, f_B(x_{m+1},...,x_n) \\ \times \prod_{j=m+1}^{n} c_{Z,j|m+1,...,j-1}[F(z|x_{m+1},...,x_{j-1}), F(x_j|x_{m+1},...,x_{j-1})] \end{aligned}$$

This is a D-Vine where Z is always an end node, and where the joint distribution of $\mathbf{X}_B$ is not subjected to a pair-copula decomposition.

Omitting the variable arguments for brevity, and assuming conditional independence for all unspecified relations, the joint density for Z, $\mathbf{X}_A$ and $\mathbf{X}_B$ is then:

$$f_{A.B,Z} = f_Z \, f_A \, f_B \prod_{i=1}^{m} c_{Z,i|1,\ldots,i-1} \prod_{j=m+1}^{n} c_{Z,j|m+1,\ldots,j-1} \tag{4}$$

This is equivalent to a regular vine formed by joining some C-vine implicit in $\mathbf{X}_A$ to some D-vine implicit in $\mathbf{X}_B$ via a supplementary node Z, given which $X_i \in \mathbf{X}_A$ and $X_j \in \mathbf{X}_B$ remain conditionally independent.

The advantage of this formulation over the industry standard approach is that it allows us to specify, and achieve through sample re-ordering alone, more than one dependence relationship between variables in $\mathbf{X}_A$ and $\mathbf{X}_B$ if we can specify these in terms of products of the form:

$$c_{Z,i} \cdot c_{Z,j} \qquad i \in A,\, j \in B$$

Products of this form are particularly tractable if all copulas are Gaussian, in which case they model $\rho_{ij} = \rho_{zi} \, \rho_{zj}$. We will consider an example of this below.

4.1 Sampling algorithm

An example of an algorithm to sample from a joint density specified in this way is as follows:

1. Define $\mathcal{V}_A$ as any C-Vine including Z and $\mathbf{X}_A = \{X_1, \ldots, X_m\}$, where Z is never the central node (alternatively, a D-vine where Z is always an end node).
   - Only the choices of the pair copulas between Z and each X_i are required, as we do not wish to alter the joint distribution of $\mathbf{X}_A$.
   - A vine structure for $\mathbf{X}_A$ can then be estimated, for instance using the approach of Dissman et al. [4].
   - The pair copulas for each edge can then be estimated, either parametrically using maximum likelihood or non-parametrically using the approach of Nagler [5].
2. Take $\mathbf{X}_A$ as an observed sample from $\mathcal{V}_A^-$ (the vine $\mathcal{V}_A$ excluding Z).
   - Using the pair copula decomposition of $\mathbf{X}_A$, find the random vector $U_{A^-} = \{U_1, \ldots, U_m\}$ where $U_i = F(X_i|X_1, \ldots, X_{i-1})$ using the Rosenblatt transformation [6].
   - Draw U_Z, a vector of observations from a standard uniform distribution, combine with U_{A^-} and let this equal U_A.
   - Using U_A, sample a vector Z_A from $\mathcal{V}_A$, using for instance the algorithm of Aas et al. [7].
3. Define $\mathcal{V}_B$ as any D-Vine including Z and $\mathbf{X}_B$, where Z is always end node (or equivalently as any C-Vine where Z is never the central node).
   - Only the choices of the pair copulas between Z and each X_i are required, as we do not wish to alter the joint distribution of $\mathbf{X}_B$.
   - A vine can be fitted to $\mathbf{X}_B$ as above.
   - Take $\mathbf{X}_B$ as observed sample from $\mathcal{V}_B^-$ (the vine $\mathcal{V}_B$ excluding Z), and determine $U_{B^-} = \{U_{m+1}, \ldots, U_n\}$ where $U_i = F(X_i|X_{m+1}, \ldots, X_{i-1})$.

- Combine U_Z with U_{B^-} and let this equal U_B.
- Using U_B, sample vector Z_B from $\mathcal{V}_B$.

4. Re-order the ranks of Z_A such that they equal the ranks of Z_B.
5. Re-order $\mathbf{X}_A$ to be consistent with Z_A, and $\mathbf{X}_B$ to be consistent with Z_B.
6. Hence the joint sample has the joint density required.

In practical terms, correlation or rank correlation targets between the sets of pre-simulated data can then be achieved by the selection of the Z-copulas, appropriately parameterised.

5 Example comparing industry standard and vine approaches

To demonstrate the proposed method and how it can improve over the industry standard approach, this section sets out a low-dimensional example, in which all variables are Normally distributed and all copulas are Gaussian. Although an assumption of multivariate Normality is not very often used for modelling insurance risks or indeed market risk variables, I use it here in any case, as an exact solution is available. In the general case, numerical approaches must be taken.

Let $\mathbf{X}_A = \{X_1, X_2, X_3\}$ be a sample from our internal model, with trivariate Normal distribution, where the variables represent reserve risk, premium risk and catastrophe risk respectively, and with correlation matrix:

$$R_A = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ 0.67 & 1 & \rho_{23} \\ 0.25 & 0.50 & 1 \end{pmatrix}$$

and let $\mathbf{X}_B = \{X_4, X_5, X_6\}$ be a sample from an external ESG, again with trivariate Normal distribution, where the variables represent equity risk, interest rate and inflation rate respectively with correlation matrix:

$$R_B = \begin{pmatrix} 1 & \rho_{45} & \rho_{46} \\ 0.30 & 1 & \rho_{56} \\ 0.15 & 0.30 & 1 \end{pmatrix}$$

Note that in each case, the variables within the internal model and within the external model have positive correlations with each other.

Say we believe that our insurance risk results are likely to be worse when the economy is performing badly—this is a common assumption for insurers writing certain financial lines of business for example, where periods of economic stress can result in reserve deterioration.

To model this, we therefore choose to associate 'bad' (low) values of the equity index variable X_4 with 'bad' (high) values of reserve risk variable X_1. One way for us to do this is by assuming these variables are negatively correlated somehow. For the sake of the example in this paper we will target a correlation of -50% between X_1 and X_4.

Recall that the 'catch', as it were, is that we can only achieve this dependence by re-ordering the rows in the two samples relative to one another, while maintaining the joint structure within each sample.

5.1 Industry standard approach

Let copula C_{14} be a bivariate Gaussian copula with correlation $\rho_{14} = -50\%$ between X_1 and X_4. Because C_{14} is Gaussian, the resulting joint distribution for all the variables is multivariate Normal.

Assuming conditional independence, the joint density can be written as:

$$f_{AB}(x_1, \ldots, x_6) = f_A(x_1, x_2, x_3)\, f_B(x_4, x_5, x_6)\, c_{14}[F_1(x_1), F_4(x_4)] \tag{5}$$

We can use (2) to find the unspecified entries to complete the 6×6 correlation matrix:

$$R_{AB} = \begin{pmatrix} R_A & Y \\ Y^T & R_B \end{pmatrix} = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} & \rho_{14}\rho_{45} & \rho_{14}\rho_{46} \\ 0.67 & 1 & \rho_{23} & \rho_{12}\rho_{14} & \rho_{12}\rho_{14}\rho_{45} & \rho_{12}\rho_{14}\rho_{46} \\ 0.25 & 0.50 & 1 & \rho_{13}\rho_{14} & \rho_{13}\rho_{14}\rho_{45} & \rho_{13}\rho_{14}\rho_{46} \\ -0.50 & -0.34 & -0.13 & 1 & \rho_{45} & \rho_{46} \\ -0.15 & -0.10 & -0.04 & 0.30 & 1 & \rho_{56} \\ -0.08 & -0.05 & -0.15 & 0.15 & 0.30 & 1 \end{pmatrix} \tag{6}$$

and see that the selection of ρ_{14} determines all the resulting correlations between the other variables in $\mathbf{X}_A$ and $\mathbf{X}_B$. So while we have achieved the target of -50% between X_1 and X_4 we also have negative correlations for every other inter-model pair.

In particular each of the correlations between the insurance risks and inflation $\{\rho_{16}, \rho_{26}, \rho_{36}\}$ are now also negative. This means we have associated low (good) values of inflation with high (bad) values of insurance risk. This is an unintended and undesirable outcome, as we typically expect high inflation to contribute to worse insurance risk outcomes, not the other way around. However, as we only have a single parameter, we cannot do anything about this.

Another way to conceptualise the problem is to consider Figure 3, the adjacency graph of R_{AB} excluding the diagonal and all conditionally independent relationships. Simply put, we would like to be able to add additional edges to this graph to control the extra dependencies we desire. However, we have the additional requirement that we can only achieve our targets by re-ordering the samples, so although there are methods to add further edges to the adjacency graph and determine the range of values which result in a valid correlation matrix, we lack a sampling algorithm.

As we will now see, the vine method allows us to achieve our aim, not only by allowing us to easily find valid parameterisations, but also because the sample algorithm relies only on re-ordering.

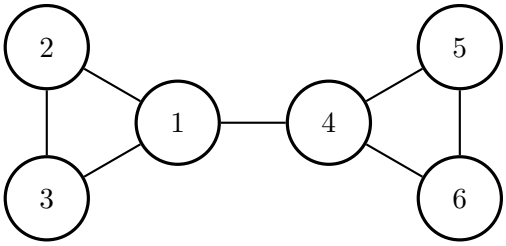

Figure 3: Adjacency graph of R_{AB}.

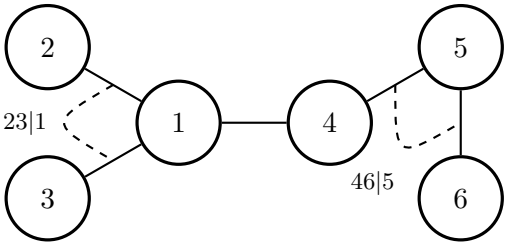

Figure 4: Vine copula on R_{AB}.

5.2 Vine copula method

Using a pair copula decomposition for the joint distribution, we can re-write (5) in the following (non-unique) way, omitting arguments:

$$\begin{aligned} f_{AB} &= f_A\, f_B\, c_{14} \\ &= [f_1\, f_2\, f_3\, c_{12}\, c_{13}\, c_{23|1}] \times [f_4\, f_5\, f_6\, c_{45}\, c_{56}\, c_{46|5}] \times c_{14} \\ &= c_{12}\, c_{13}\, c_{14}\, c_{45}\, c_{56}\, c_{23|1}\, c_{46|5} \prod_{i=1}^{6} f_i \end{aligned} \tag{7}$$

This factorisation of the joint density is expressed as a vine in Figure 4 where the nodes represent variables, solid edges between nodes represent the unconditional Gaussian pair copulas, and dashed lines represent the conditional pair copulas $C_{23|1}$ and $C_{46|5}$, which are implicit in $\mathbf{X}_A$ and $\mathbf{X}_B$ respectively. The similarities with Figure 3 are clear.

Because $\mathbf{X}_A$ and $\mathbf{X}_B$ are Normally distributed, we have the nice property that the conditional copulas are also Gaussian, with conditional correlations $\rho_{23|1}$ and $\rho_{46|5}$. Further, as shown in [8], for any bivariate distribution with linear regression, the partial and conditional correlations are equal. As the Normal distribution has this property, from the definition of partial correlation $\rho_{ij;k}$ we have the following relationship:

$$\rho_{ij|k} = \rho_{ij;k} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{\sqrt{1-\rho_{ik}^2}\sqrt{1-\rho_{jk}^2}} \tag{8}$$

This gives us a convenient way to convert between the conditional correlations which parameterise the conditional copulae, and the entries of the correlation matrix. As we can now associate a partial correlation with each edge in the vine, this is a so-called partial correlation vine.

Recall from (6) that when using the industry standard approach to model the desired negative correlation between reserve risk and equity market index, we necessarily ended up with unwanted negative correlations between the inflation rate and each of the insurance risks. We can now apply our main result from (4) by including our supplementary variable Z to avoid this outcome.

Let the joint density of $\mathbf{X}_A$, $\mathbf{X}_B$ and Z be:

$$f_{ABZ}(x_1, \ldots, x_6, z) = f_A\, f_B\, f_Z\, c_{Z,1}\, c_{Z,2|1}\, c_{Z,3|12}\, c_{Z,4}\, c_{Z,5|4}\, c_{Z,6|45} \tag{9}$$

where all of the copula densities involving Z are Gaussian. Therefore the joint distribution for all variables is Normal with correlation matrix:

$$R_{ABZ} = \begin{pmatrix} R_A & Z_A & Y \\ Z_A^T & 1 & Z_B^T \\ Y^T & Z_B & R_B \end{pmatrix} \qquad \text{where:} \quad Z_A = \begin{pmatrix} \rho_{Z,1} \\ \rho_{Z,2} \\ \rho_{Z,3} \end{pmatrix} \quad \text{and} \quad Z_B = \begin{pmatrix} \rho_{Z,4} \\ \rho_{Z,5} \\ \rho_{Z,6} \end{pmatrix}$$

To find the entries of Z_A and Z_B, we can recursively apply the definition for partial correlation to find the correlation between the specified pairs given the parameters for the copulas in the vine, for example:

$$\rho_{Z,2} = \rho_{Z,2|1}\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{Z,1}^2} + \rho_{12}\,\rho_{Z,1}$$

To find entries of Y, recall that all $\{X_i, X_j\}, i \in A, jB$ are conditionally independent, i.e. $\rho_{ij|Z} = 0$. Hence we find:

$$Y = Z_A Z_B^T = \begin{pmatrix} \rho_{Z,1}\rho_{Z,4} & \rho_{Z,1}\rho_{Z,5} & \rho_{Z,1}\rho_{Z,6} \\ \rho_{Z,2}\rho_{Z,4} & \rho_{Z,2}\rho_{Z,5} & \rho_{Z,2}\rho_{Z,6} \\ \rho_{Z,3}\rho_{Z,4} & \rho_{Z,3}\rho_{Z,5} & \rho_{Z,3}\rho_{Z,6} \end{pmatrix}$$

Compare this result to the industry standard approach. We now have six parameters to control the inter-model correlations and can specify more than one target. To resolve the issue of negative correlation between inflation and insurance risks, we specify that the correlations between each insurance risk and inflation should be no less than 5%. We therefore have the following requirements:

$$\begin{aligned} \rho_{14} &= \rho_{Z,1}\rho_{Z,4} = -0.5 \\ \rho_{i6} &= \rho_{Z,i}\rho_{Z,6} \geq 0.05 \, , i \in \{1,2,3\} \end{aligned} \tag{10}$$

This leaves us with several unspecified entries in Y. To find a unique solution we therefore include another requirement: that we seek the solution which introduces the minimum amount of extra information beyond our specified targets. For the Normal distribution this is equivalent to maximising the determinant of the correlation matrix [9], which corresponds to maximising entropy.

We parameterise the vine copula by vector $\theta^T = (\rho_{Z,1}, \rho_{Z,2|1}, \rho_{Z,3|12}, \rho_{Z,4}, \rho_{Z,5|4}, \rho_{Z,6|54})$, where each element represents the (un)conditional correlation linking Z to the respective X_i. Clearly each value is bounded in the range $[-1, 1]$. Subject to our constraints (10), the optimal parameter vector θ^* is found by solving the following optimization problem:

$$\theta^* = \arg\max_{\theta} \det R_{ABZ}$$

Although maximising the determinant of R_{ABZ} directly is possible, we can leverage the partial correlation vine structure for R_{AB} from (7) to rewrite (9) as a partial correlation vine for R_{ABZ}, as shown in Figure 5. We can then rely on the following result: the determinant D of an n-dimensional correlation matrix R, can be expressed as the following product of terms involving the partial correlations and edges from $\mathcal{V}$, a partial correlation vine on R [10]:

$$D = \prod_{e \in E(\mathcal{V})} (1 - \rho_{e_1,e_2;D_e})^2$$

The benefit of this approach is that clearly we can simplify the optimisation problem by setting to zero the partial correlation for any edges in the graph where we have assumed conditional independence. We find the following solution for θ^* and Y^*, the matrix of achieved inter-model correlations:

$$\theta^* = \begin{pmatrix} 0.72 \\ 0.08 \\ 0.18 \\ -0.69 \\ 0.10 \\ 0.36 \end{pmatrix} \qquad Y^* = \begin{pmatrix} -0.50 & -0.10 & 0.11 \\ -0.36 & -0.07 & 0.08 \\ -0.21 & -0.04 & 0.05 \end{pmatrix}$$

We can now see that in addition to achieving the target correlation of -50% between reserve risk X_1 and equity index X_4, using the vine copula approach we can also achieve an additional target,

of at least 5% correlation between each of reserve risk X_1, premium risk X_2, catastrophe risk X_3 and inflation rate X_6.

Given θ^* and pre-simulated samples $\mathbf{X}_A$, $\mathbf{X}_B$ one can now produce a sample from the joint distribution which has the inter-model correlations Y^*, using the algorithm described earlier.

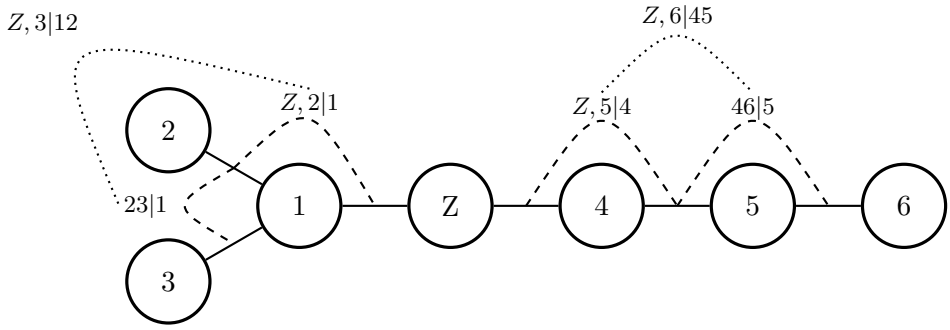

Figure 5: Vine copula on R_{ABZ}.

6 Conclusion

This short paper describes a problem often faced in capital modelling—that of specifying and achieving dependencies between internal and external models—and provides a method which goes some way to addressing the limitations of the industry standard approach. It does this by allowing for the specification of more than one dependence relationship between sets of pre-simulated data, which can be used either to achieve more than one target correlation between pairs, or just to control for unintended and unwanted side-effects of targeting a specific important correlation, as in the example.

References

[1] Sklar, M., 1959. *Fonctions de répartition à n dimensions et leurs marges.* Annales de l'ISUP, 8(3), pp.229–231.

[2] Georgescu, D.I., Higham, N.J. and Peters, G.W., 2018. *Explicit solutions to correlation matrix completion problems, with an application to risk management and insurance.* Royal Society open science, 5(3), p.172348.

[3] Bedford, T. and Cooke, R.M., 2002. *Vines: A New Graphical Model for Dependence.* The Annals of statistics, 30(4), pp.1031-1068.

[4] Dissmann, J., Brechmann, E.C., Czado, C. and Kurowicka, D., 2013. *Selecting and estimating regular vine copulae and application to financial returns.* Computational Statistics & Data Analysis, 59, pp.52-69.

[5] Nagler, T., 2018. *kdecopula: An R package for the kernel estimation of bivariate copula densities.* Journal of Statistical Software, 84, pp.1-22.

[6] Rosenblatt, M., 1952. *Remarks on a multivariate transformation.* The annals of mathematical statistics, 23(3), pp.470-472.

[7] Aas, K., Czado, C., Frigessi, A. and Bakken, H., 2009. *Pair-copula constructions of multiple dependence*. Insurance: Mathematics and economics, 44(2), pp.182-198.

[8] Kurowicka, D. and Cooke, R., 2000. *Conditional and partial correlation for graphical uncertainty models*. Recent Advances in Reliability Theory: Methodology, Practice, and Inference (pp. 259-276). Boston, MA: Birkhäuser Boston.

[9] Whittaker, J., 2009. *Graphical Models in applied multivariate statistics*, Wiley Publishing.

[10] Kurowicka, D. and Cooke, R.M., 2006. *Completion problem with partial correlation vines*. Linear algebra and its applications, 418(1), pp.188-200.