

Approximate Uncertainty Modeling in Risk Analysis with Vine Copulas

Tim Bedford,¹ Alireza Daneshkhah,² and Kevin J. Wilson^{1,*}

Many applications of risk analysis require us to jointly model multiple uncertain quantities. Bayesian networks and copulas are two common approaches to modeling joint uncertainties with probability distributions. This article focuses on new methodologies for copulas by developing work of Cooke, Bedford, Kurowica, and others on vines as a way of constructing higher dimensional distributions that do not suffer from some of the restrictions of alternatives such as the multivariate Gaussian copula. The article provides a fundamental approximation result, demonstrating that we can approximate any density as closely as we like using vines. It further operationalizes this result by showing how minimum information copulas can be used to provide parametric classes of copulas that have such good levels of approximation. We extend previous approaches using vines by considering nonconstant conditional dependencies, which are particularly relevant in financial risk modeling. We discuss how such models may be quantified, in terms of expert judgment or by fitting data, and illustrate the approach by modeling two financial data sets.

KEY WORDS: Copula; entropy; information; risk modeling; vine

1. INTRODUCTION

Many areas of applied risk analysis require us to model multiple uncertainties using multivariate distributions. For some decision support settings, it is common to use discrete models such as Bayesian networks. In other settings, particularly when modeling financial data or carrying out uncertainty analysis, it is necessary to have models of multivariate continuous random variables. Dependency modeling is therefore an area of great interest for a whole range of risk analysis applications.

There is a growing literature on the use of copulas to model dependencies (see, e.g., surveys by

Nelsen and^(1,2) Joe⁽³⁾). Copulas have found application in a number of areas, including combining expert opinion and stochastic simulation.⁽⁴⁻⁹⁾ A copula is a joint distribution on the unit square (or more generally on the unit n -cube) with uniform marginal distributions. Under reasonable conditions, we can uniquely specify a joint distribution for n random variables by specifying the univariate distribution for each variable, and, in addition, specifying the copula. This is because we can simply transform each variable by its own distribution function (sometimes called its quantile function) to ensure that the transformed variable has a uniform distribution, so that the joint distribution function F can be written:

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)), \quad (1)$$

where C is a copula distribution function, and F_1, \dots, F_n are the univariate, or marginal, distribution functions. We can use this formula constructively: given a copula C and marginals F_1, \dots, F_n we can *define* F in this way. A special case is that

¹Department of Management Science, University of Strathclyde, Glasgow, UK.

²Cranfield Water Science Institute, Cranfield University, Bedford, UK.

*Address correspondence to Dr. Kevin Wilson, Department of Management Science, University of Strathclyde, Glasgow G1 1XQ, UK; kevin.j.wilson@strath.ac.uk.

of the “Gaussian copula,” obtained from the Gaussian joint distribution and parameterized by the correlation matrix. Use of the Gaussian copula to construct joint distributions is equivalent to the NORTA method (normal to anything).⁽¹⁰⁾

The use of a copula to model dependency is simply a translation of one difficult problem into another: instead of the difficulty of specifying the full joint distribution we have the difficulty of specifying the copula. The main advantage is the technical one that copulas are normalized to have support on the unit square and uniform marginals. As many authors restrict the copulas to a particular parametric class (Gaussian, multivariate t , etc.) the potential flexibility of the copula approach is not realized in practice. The approach used in this article, by contrast, allows a lot of flexibility in copula specification. It utilizes a graphical model, called a vine, to systematically specify how two-dimensional copulas are stacked together to produce an n -dimensional copula.

The main objectives of this article are to show that a vine structure can be used to approximate any given multivariate copula to any required degree of approximation, and to show how this can be operationalized for use in practical situations involving uncertain risks. The standing technical assumptions we make are that the multivariate copula density f under study is continuous and is nonzero. No other assumptions are needed. We illustrate this by modeling a data set of Norwegian financial data that was previously analyzed in Aas *et al.*⁽¹¹⁾ We extend the modeling approach used by Aas *et al.*⁽¹¹⁾ by considering the possibility of nonconstant conditional dependencies within the vine structure.

Since vines demonstrate high flexibility and advantages in constructing multivariate distributions, they have recently been used to describe the interdependence structure and build the joint distribution of portfolio returns. As coherent measures of risk, value at risk (VaR) and conditional value at risk (CVaR), which are greatly affected by the tail distribution of risk factors, have been widely used to optimize portfolios and measure their risk. Deng *et al.*⁽¹²⁾ used extreme value theory to model the tails of the innovation of each asset return and estimate risk of assets. The dependence structure between innovations of asset returns can be represented by a vine. This vine is useful to model both the influence of portfolio dimensions and the differences of tail dependence between assets. As expected, the optimal portfolio is better via a vine than that via a Student copula model (see also Ref. 11 for similar study). We

illustrate that the minimum information vine can outperform the standard multivariate copula model and specific parametric vines.⁽¹³⁾

Our constructive approach involves the use of minimum information copulas that can be specified to any required degree of precision based on the data available. We prove rigorously that good approximation “locally” guarantees good approximation globally. Finally, we discuss rules of thumb that could be used to apply this in practice. In particular, we discuss vine structure. A vine structure imposes no restrictions on the underlying joint probability distribution it represents (as opposed to the situation for Bayesian networks, for example). However, this does not mean that we should ignore the question about which vine structure is most appropriate, for some structures allow the use of less complex conditional copulas than others. Conversely, if we only allow certain families of copulas, then one vine structure might fit better than another.

2. VINE CONSTRUCTIONS FOR MULTIVARIATE DEPENDENCE

A copula is a multivariate distribution function with standard uniform marginal distributions. Using Equation (1), a copula can be used, in conjunction with the marginal distributions, to model any multivariate distribution. However, apart from the multivariate Gaussian, Student, and the exchangeable multivariate Archimedean copulas, the set of higher dimensional copulas proposed in the literature is limited and is not rich enough to model all possible mutual dependencies among the n variates (see Kurowicka and Cooke⁽¹⁴⁾ for details of these copulas). Hence, it is necessary to consider more flexible constructions.

A structure, here denoted the *pair-copula construction* or *vine*, allows for the free specification of (at least) $n(n-1)/2$ copulas between n variables. (Note that $n(n-1)/2$ is the number of entries above the diagonal of an $n \times n$ correlation matrix—though these are algebraically related so not completely free variables.) This structure was originally proposed by Joe,⁽³⁾ and reformulated and discussed in detail by Bedford and Cooke,^(13,15) who considered simulation, information properties, and the relationship to the multivariate normal distribution but who considered a more general method called a Cantor tree construction.

Kurowicka and Cooke⁽¹⁴⁾ (Chapters 4, 6–9) consider simulation issues and Aas *et al.*⁽¹¹⁾ look at

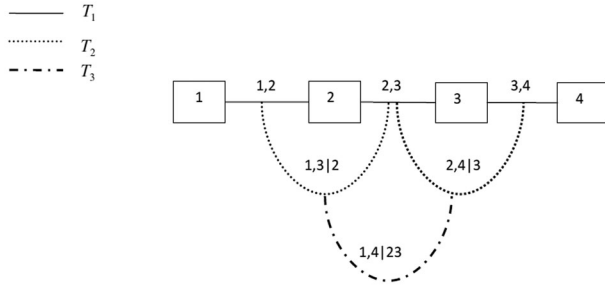


Fig. 1. A regular vine with four elements.

inference. Excellent overviews of vines are given in Refs. 16 and 17. The modeling scheme is based on a decomposition of a multivariate density into a set of bivariate copulas. The way these copulas are built up to give the overall joint distribution is determined through a structure called a vine, and can be easily visualized. A vine on n variables is a nested set of trees, where the edges of the tree j are the nodes of the tree $j + 1$ (for $j = 1, \dots, n - 2$), and each tree has the maximum number of edges. For example, Fig. 1 shows a vine with four variables, which consists of three trees (T_1, T_2, T_3) with 3, 2, and 1 edges, respectively. A *regular vine* on n variables is a vine in which two edges in tree j are joined by an edge in tree $j + 1$ only if these edges share a common node, for $j = 1, \dots, n - 2$. There are $n(n - 1)/2$ edges in a regular vine on n variables. The formal definition is as follows.

DEFINITION 1. (Vine, regular vine) \mathcal{V} is a vine on n elements if

- (1) $\mathcal{V} = (T_1, \dots, T_{n-1})$.
- (2) T_1 is a connected tree with nodes $N_1 = \{1, \dots, n\}$ and edges E_1 ; for $i = 2, \dots, n - 1$, T_i is a connected tree with nodes $N_i = E_{i-1}$.

\mathcal{V} is a regular vine on n elements if additionally the proximity condition holds.

- (3) For $i = 2, \dots, n - 1$, if a and b are nodes of T_i connected by an edge in T_i , where $a = \{a_1, a_2\}$, $b = \{b_1, b_2\}$, then exactly one of the a_i equals one of the b_i .

One of the simplest regular vines is shown in Fig. 1—this structure is called a *D-vine*; see Kurowicka and Cooke,⁽¹⁴⁾ p. 93. Here, T_1 is the tree consisting of the straight edges between the numbered nodes, T_2 is the tree consisting of the curved edges that join the straight edges in T_1 , and so on.

For a regular vine, each edge of T_1 is labeled by two numbers from $\{1, \dots, n\}$. If we take two edges of

T_1 , for example, 12 and 23, which are nodes joined by an edge in T_2 , then of the numbers labeling these edges one is common to both (2), and they both have one unique number (1,3, respectively). The common number(s) will be called the *conditioning set* D_e for that edge e (in this example, the conditioning set is simply $\{2\}$) and the other numbers will be called the *conditioned set* (in this example, $\{1, 3\}$). For a regular vine, the conditioned set always contains two elements.

We associate a vine distribution to a vine by specifying a copula to each edge of T_1 and a family of conditional copulas for the conditional variables given the conditioning variables, as shown by the following result of Bedford and Cooke.⁽¹⁵⁾

THEOREM 1. Let $\mathcal{V} = (T_1, \dots, T_{n-1})$ be a regular vine on n elements. For each edge $e(j, k) \in T_i, i = 1, \dots, n - 1$ with conditioned set $\{j, k\}$ and conditioning set D_e , let the conditional copula and copula density be $C_{j|k|D_e}$ and $c_{j|k|D_e}$, respectively. Let the marginal distributions F_i with densities $f_i, i = 1, \dots, n$ be given. Then, the vine-dependent distribution is uniquely determined and has a density given by:

$$f(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i) \prod_{j=1}^{n-1} \prod_{e(j,k) \in E_i} c_{j|k|D_e}(F_{j|D_e}(x_j), F_{k|D_e}(x_k)). \quad (2)$$

Note that we use $c_{j|k|D_e}$ here to be a conditional copula density and not the more usual conditional bivariate cumulative distribution function (cdf) (which is not a copula). The existence of regular vine distributions is discussed in detail by Bedford and Cooke.⁽¹³⁾

The density decomposition associated with four random variables $\mathbf{X} = (X_1, \dots, X_4)$ with a joint density function $f(x_1, \dots, x_4)$ satisfying a copula-vine structure shown in Fig. 1 with the marginal densities f_1, \dots, f_4 is:

$$\begin{aligned} f_{1234}(x_1, \dots, x_4) &= \prod_{i=1}^4 f(x_i) \times c_{12}(F(x_1), F(x_2)) \\ &\quad c_{23}(F(x_2), F(x_3)) c_{34}(F(x_3), F(x_4)) \times \\ &\quad c_{13|2}(F(x_1 | x_2), F(x_3 | x_2)) c_{24|3}(F(x_2 | x_3), \\ &\quad F(x_4 | x_3)) \times c_{14|23}(F(x_1 | x_2, x_3), F(x_4 | x_2, x_3)). \end{aligned} \quad (3)$$

Note that in the special case of a joint normal distribution, we would use the normal copula everywhere in the above expression and the conditional copulas would be constant (i.e., not depend on the conditioning variable). This means that the joint normal structure is specified by $n(n-1)/2$ (conditional) correlation values, which are algebraically free between -1 and $+1$ (unlike the values in a correlation matrix). See Bedford and Cooke⁽¹³⁾ for more details.

Theorem 1 gives us a constructive approach to build a multivariate distribution given a vine structure: if we choose marginal densities and copulas, then this will give us a multivariate density. Hence, vines can be used to model general multivariate densities. However, in practice we have to use copulas from a convenient class, and this class should ideally be one that allows us to approximate any copula to an arbitrary degree. In the following sections, we address this issue in more detail. By having this class of copulas, we can approximate any multivariate distribution using any vine structure.

Unlike the situation with Bayesian networks, where not all structures can be used to model a given distribution, the theorem shows that, in principle, any vine structure may be used to model a given distribution. However, when specific families of copulas are used some vine structures work better than others. That is, given a family of copulas, some vine structures give a better degree of approximation than others. We shall return to this point later.

Much work has been done to operationalize the use of vines for modeling multivariate data sets^(11,16,18,19) and an R package “VineCopula” has been developed to implement the approaches in this work (<http://cran.r-project.org/web/packages/VineCopula/index.html>).

It is worth stressing that the flexibility of vines gives us the potential to capture any fine-grain structure within a multivariate distribution. A key aspect that cannot be modeled by Bayesian networks is that of conditional dependence. Bayesian networks are built around the concept of conditional *independence*—arrows from a parent node to two child nodes means that the child variables are conditionally, independent given the parent variable. Of course, unconditionally these two child nodes are dependent. However, different models of conditional dependence are not available as building blocks in Bayesian networks.

Multivariate Gaussian copulas do allow for a specification of conditional dependence, but do not

allow that dependence to change: in a multivariate normal distribution, the conditional correlation of two variables given a third may be nonzero but is always constant. Our approach allows the explicit modeling of nonconstant conditional dependence, as we illustrate with a simple example.

The deeper a bivariate copula is in the vine hierarchy, the more variables will be conditioned on. If the conditional dependencies are neglected, then vines are a direct method to build flexible multivariate models using bivariate copulas as building blocks. Acar *et al.*⁽²⁰⁾ argue that ignoring conditional dependencies (the so-called simplifying assumption) can lead to reasonably precise approximations of the underlying copula (as claimed by Ref. 21), but this can in general be misleading. They develop an approach to condition parametric bivariate copulas on a single scalar variable. Stoeber *et al.*⁽²²⁾ repeated this concern, after studying several examples, and felt the assumption of an absolutely continuous vine is sometimes too strong. The latter assumption is used to make the vine models tractable for inference and model selection. Lopez-Paz *et al.*⁽²³⁾ also reported that the simplifying assumption can lead to oversimplified estimates in practice. They then extended the work of Acar *et al.* by developing a method for estimation of fully conditional vines using a Gaussian Process.

2.1. Example

We consider an example involving nonconstant conditional correlations. Suppose we have three unknown quantities, X_1, X_2, X_3 , for which we wish to specify a joint distribution. Marginally each variable is normally distributed, $X_i \sim N(m_i, s_i^2)$, for $i = 1, 2, 3$, and X_i is not independent of X_j for $i \neq j$. We can represent the joint distribution between X_1, X_2, X_3 using a D-vine in three dimensions. That is, specify a copula between X_1, X_2 , one between X_2, X_3 , and then a conditional copula between $X_1, X_3 | X_2$.

In each case, we choose a bivariate Gaussian copula. This takes the form, for $U_i = F(x_i)$,

$$C(u_i, u_j) = \Phi_\rho(\Phi^{-1}(u_i), \Phi^{-1}(u_j)),$$

where $\Phi(\cdot)$ is the cdf of the standard Normal distribution and $\Phi_\rho(\cdot, \cdot)$ is the cdf of the standard bivariate Normal distribution with correlation ρ . Suppose that the correlations in the first tree of the vine are specified as ρ_{12} and ρ_{23} for the marginal copulas between X_1, X_2 and X_2, X_3 , respectively.

If we were to specify a constant correlation between $X_1, X_3 \mid X_2$ then the resulting distribution of X_1, X_2, X_3 could be modeled using the Gaussian copula. However, let us suppose that the correlation between $X_1, X_3 \mid X_2$ is not constant but rather

$$\begin{aligned}\rho_{X_1, X_3 \mid X_2 \in (0; 0.33)} &= 1, \quad \rho_{X_1, X_3 \mid X_2 \in (0.33; 0.67)} = 0, \\ \rho_{X_1, X_3 \mid X_2 \in (0.67; 1)} &= -1,\end{aligned}$$

so that there is a positive linear relationship between the variables for $U_2 = F_2(X_2)$ in $(0, \frac{1}{3})$, they are *uncorrelated* for U_2 between $(\frac{1}{3}, \frac{2}{3})$, and there is a negative linear relationship between them in $(\frac{2}{3}, 1)$.

We can divide the support of X_2 , via U_2 , into intervals. Then we can define a Gaussian copula within each interval. Suppose that numerical values for the required means and standard deviations are $m_1 = 0.5, m_2 = 1, m_3 = -1$, and $s_1 = s_2 = s_3 = 2$ and the correlations between X_1, X_2 and X_2, X_3 are $\rho_{12} = 0.75, \rho_{23} = -0.75$, respectively. This fully specifies the vine.

We can simulate from the vine to check that we recover the conditional correlations for $X_1, X_3 \mid X_2$. In order to do this, we randomly draw u_1, u_2, u_3 , three standard uniform variables. Then

$$x_1 = F_1^{-1}(u_1), x_2 = F_{2|1}^{-1}(u_2 \mid x_1), x_3 = F_{3|12}^{-1}(u_3 \mid x_1, x_2),$$

where the distribution function $F_{3|12}$ is found from $f_{3|1,2}(x_3 \mid x_1, x_2)$. For further details on this, see Ref. 15 and Section 5.2.1. We perform 5,000 simulations. The resulting X_1, X_3 values are plotted in Fig. 2.

We have recovered the correlations well in each interval. The simulated correlations are $\rho_{X_1, X_3 \mid U_2 \in (0; 0.33)} = 0.9998$, $\rho_{X_1, X_3 \mid U_2 \in (0.33; 0.67)} = -0.0032$, and $\rho_{X_1, X_3 \mid U_2 \in (0.67; 1)} = -0.9998$.

We shall use this approach when considering the example of Aas *et al.*⁽¹¹⁾ in Section 5.1. We believe that incorporating such nonconstant conditional correlations as we lay out in this article would be a useful addition to the R package VineCopula. By using a smooth function for the dependence instead of a piece-wise constant function as in the binning in the example above it, would be possible to apply such an approach to large, complex distributions.

The use of Gaussian copulas in financial modeling has come under fire for its uncritical use. Shreve⁽²⁴⁾ points out that the simple modeling of correlation available in the Gaussian copula does not pass validation tests, but that this did not stop its widespread adoption in the finance community.

3. BUILDING BIVARIATE MINIMUM INFORMATION COPULAS

The emphasis in this article is on approximation rather than on statistically optimal estimation techniques. We use minimum information methods to operationalize the approximation in the class of copulas used. This section discusses how the data required to specify bivariate copulas can be derived, either from expert or sampling data, and shows how this can be used to determine a minimum information copula.

We recall that when f and g are bivariate densities then the *relative information* of f with respect to g is:

$$I(f|g) = \int \int \ln(f(x, y)/g(x, y)) f(x, y) dx dy.$$

Information is a measure of the degree of deviation of f from g and is minimized at 0 when $f = g$. Furthermore, because the information function is transformation-invariant, the relative information of f with respect to g is the same as that between the copula for f with respect to the copula of g . This makes information a natural quantity with which to measure the degree of dependency in a copula, for if g is an independent bivariate with the same marginal distributions as f , then $I(f|g)$ is the same as the information of the copula of f relative to the independent copula.

From the perspective of information and entropy,⁽²⁵⁾ a natural way to specify dependency constraints is through the use of moments. These can be specified either on the copula or on the underlying bivariate density (as long as we know the marginal distributions and can therefore transform from one to the other). We consider moment constraints in which real-valued functions h_1, \dots, h_k are required to take expected values $\alpha_1, \dots, \alpha_k$. By a minimum information copula, we mean a copula that satisfies a set of constraints as above and that has minimum information (with respect to the uniform copula $c(u, v) = uv$) among the class of all copulas satisfying those constraints. This copula (when it exists and is unique—which is normally the case) is the “most independent” bivariate density that meets the constraints. Note that probabilities are simply expectations of identity functions and so this method of specifying constraints is not restrictive.

Information has a further advantage for us in that it is a natural measure for vine distributions: a specification of minimum information bivariate copulas automatically gives minimum information vine

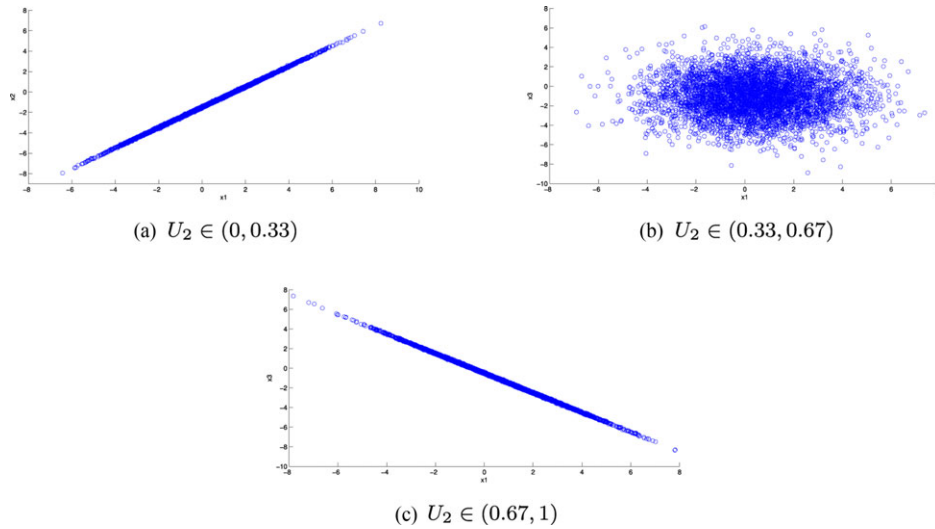


Fig. 2. The simulated distributions of X_1 , X_3 given X_2 in each of the intervals.

distributions. Specifically, lemma 4.4 and theorem 4.5 in Bedford and Cooke⁽¹³⁾ (see also Kurowicka and Cooke⁽¹⁴⁾) show that if we take a minimal information copula satisfying each of the (local) constraints (on moments, rank correlation, etc.), then the resulting joint distribution is also minimally informative given those constraints.

3.1. Data: Expert Judgment or Random-Sample-wBased Approaches

Quantitative operations research models are typically quantified either by expert judgment or estimation from data. In our case, the minimum information models are parameterized by the expected values of functions $h_i : [0, 1]^2 \rightarrow \mathbb{R}$ discussed above. The simplest case is to consider a single function defined on the copula parameters $h(u, v) = uv$. Specifying the expected value of this is equivalent to specifying the Spearman rank correlation coefficient for the copula.⁽²⁶⁾ If we wanted to consider the product-moment correlation, this would entail transforming back to the original variables and using the function:

$$h(u, v) = F_1^{-1}(u)F_2^{-1}(v),$$

where F_1 and F_2 are the marginal distributions of the original variables. The use of experts to specify correlations has been explored extensively in the literature (see, for example, Clemen and Reilly⁽⁷⁾). Hence, the methods we propose allow for common correlation-based approaches to specifying depen-

dence, as well as providing for a wider range of constraints if desired. Kurowicka *et al.*⁽²⁷⁾ explored the use of Bayesian networks to structure the specification of parameters for vine models.

We remark that the Spearman correlation can take any value between -1 and $+1$, whereas the product-moment correlation is typically restricted to a narrower interval depending on the marginal distributions involved. Bedford⁽²⁸⁾ discussed the possibilities of using the minimum information approach to explore the range of feasible values to aid experts in choosing consistent parameter values.

The approach taken in this context is subjectivist and follows a tradition in which expectation values are used to specify uncertain quantities.^(25,29–31) Within a conventional Bayesian approach, our work may be thought of as a way to generate an informative prior distribution. We are not suggesting that the approach be used as an alternative to Bayesian updating. We remark that MCMC methods have been used in conjunction with vines^(16,32) in order to update vines.

The elicitation of a joint probability distribution from experts or the approximation of a joint distribution of multiple uncertain quantities are among the key research areas in risk assessment, and the distinction between sources of uncertainty often comes into play in the elicitation of the uncertain quantities.^(33,34) Uncertainties are sometimes distinguished as being either aleatory (stochastic) or epistemic. The former arises because of natural, unpredictable variation in the performance of the system under study. In this case, the proposed

method in this article can be used to approximate the joint distribution based on observed sample data for multiple uncertain quantities. Epistemic uncertainty is due to a lack of knowledge about the behavior of the system. This is conceptually resolvable.

The epistemic uncertainty can, in principle, be eliminated with sufficient study. Borgonovo⁽³⁵⁾ and Aven⁽³³⁾ reported that subjective probabilities are often used for representing this type of uncertainty, but several other approaches can be used to represent this uncertainty. Therefore, our method can be used to elicit the prior distribution of unknown parameters by building a subjective multivariate distribution based on observable quantities. Although one may use rank correlations that are not observable quantities, within a minimum information framework it is possible to specify the expected value of any particular function on the probability space. Rank correlation falls into this framework as it is linearly related to the expected value of a product of cdfs in the copula space.

If we wish to fit distributions on the basis of sampling data (large quantities of which may be available, for example, in financial risk modeling problems), the data can be transformed to uniform after estimation of the marginals. This makes it possible to consider approximation, or encoding, of the data using a multivariate copula, and enables us to consider ways of judging how well that approximation can be made using given families of two-dimensional copulas. We give examples later in the article to illustrate this approach.

3.2. The D_1AD_2 Algorithm and Minimum Information Copulas

Suppose there are k functions, $h_1, h_2, \dots, h_k : [0, 1]^2 \rightarrow \mathbb{R}$, for which we specify the mean values $\alpha_1, \dots, \alpha_k$ that these functions simultaneously take. Further suppose that h_i, h_j are linearly independent for $i \neq j$. We seek a copula that has these mean values, a problem that is usually either infeasible or underdetermined. Assuming feasibility for the moment, we ask that the copula be minimally informative (relative to the uniform distribution), which guarantees a unique and reasonable solution. Define the kernel:

$$A(u, v) = \exp(\lambda_1 h_1(u, v) + \dots + \lambda_k h_k(u, v)). \quad (4)$$

According to the general theory of Borwein *et al.*⁽³⁶⁾ and Nussbaum⁽³⁷⁾ (section 4), there is a unique copula with minimum information satisfying the

constraints that the mean value of h_i is α_i ($i = 1, \dots, k$), and this has density

$$d^{(1)}(u)d^{(2)}(v)A(u, v)$$

for some functions $d^{(1)}(\cdot), d^{(2)}(\cdot)$. The parameters $(\lambda_1, \dots, \lambda_k)$ depend on $(\alpha_1, \dots, \alpha_k)$ in a nonlinear way. There are numerical procedures to determine this relationship: given $(\lambda_1, \dots, \lambda_k)$ we can numerically determine the functions $d^{(1)}(u)$ and $d^{(2)}(v)$ and calculate the associated mean values for h_1, h_2, \dots, h_k . By numerically solving this function, as discussed below, we can find the unique $(\lambda_1, \dots, \lambda_k)$ for which the mean values of h_1, h_2, \dots, h_k are $\alpha_1, \dots, \alpha_k$. A summary of the theory based on Bedford and Meeuwissen,⁽²⁶⁾ Nussbaum⁽³⁷⁾ (section 4), and Borwein *et al.*⁽³⁶⁾ is addressed in Ref. 38.

The general theory says that the set of all possible expectation vectors $(\alpha_1, \dots, \alpha_k)$ that could be taken by (h_1, h_2, \dots, h_k) under some probability distribution is convex, and that for every $(\alpha_1, \dots, \alpha_k)$ in the interior of that convex set there is a density with parameters $(\lambda_1, \dots, \lambda_k)$ for which (h_1, h_2, \dots, h_k) take these expectations.

This general approach to defining a copula was used by Bedford and Meeuwissen⁽²⁶⁾ with a single function $h(u, v) = uv$, which measures the Spearman rank correlation of the copula. Bedford⁽²⁸⁾ and Lewandowski⁽³⁹⁾ have considered larger groups of functions.

The discrete version of this problem can be written in terms of matrices. In this case, the probability densities defined above are approximated by probability mass functions, which are given below. Suppose that (u, v) are discretized into n points, respectively, as u_i , and v_j , $i, j = 1, \dots, n$. Then, we write $A = (a_{ij})$, $D_1 = \text{diag}(d_1^{(1)}, \dots, d_n^{(1)})$, $D_2 = \text{diag}(d_1^{(2)}, \dots, d_n^{(2)})$, where $a_{ij} = A(u_i, v_j)$, $d_i^{(1)} = d^{(1)}(u_i)$, $d_j^{(2)} = d^{(2)}(v_j)$. The assumption of uniform marginals means that:

$$\sum_j d_i^{(1)} d_j^{(2)} a_{ij} = 1/n, \text{ and } \sum_i d_i^{(1)} d_j^{(2)} a_{ij} = 1/n,$$

$$\forall i, j = 1, \dots, n.$$

Hence,

$$d_i^{(1)} = \frac{1}{n \sum_j d_j^{(2)} a_{ij}} \text{ and } d_j^{(2)} = \frac{1}{n \sum_i d_i^{(1)} a_{ij}}.$$

Finding matrices D_1 and D_2 so that D_1AD_2 is a stochastic matrix has been long studied. Sinkhorn and Knopp⁽⁴⁰⁾ gave a simple algorithm, and the iterative proportional fitting (IPF) algorithm⁽⁴¹⁾ has been

much used. IPF uses an iterative procedure to determine the entries of D_1 and D_2 . The idea is simple: start with arbitrary positive initial matrices for D_1 and D_2 , then successively define new vectors by iterating the maps:

$$d_i^{(1)} \mapsto \frac{1}{n \sum_j d_j^{(2)} a_{ij}} \quad (i = 1, \dots, n),$$

$$d_j^{(2)} \mapsto \frac{1}{n \sum_i d_i^{(1)} a_{ij}}, \quad (j = 1, \dots, n).$$

This iteration converges geometrically to give us the vectors required. Nussbaum⁽³⁷⁾ (section 4) considered the problem in greater generality, considering continuous densities and functions, and showed that the corresponding functional is a contraction mapping on a space of functions endowed with a Hilbert projective metric. We make use of this fact when considering the quality of approximations made to copulas below.

To compare different discretizations (for different n), we multiply each cell weight $d_i^{(1)} d_j^{(2)} a_{ij}$ by n^2 as this quantity approximates the continuous copula density with respect to the uniform distribution.

As discussed above, for a given set of functions (h_1, \dots, h_k) , the mapping from the set of vectors of λ s parameterizing the kernel A onto the expectations of the function $(\alpha_1, \dots, \alpha_k)$ is found numerically, and optimization techniques are used. We wish to determine the appropriate set of λ s for given expectations α_i , where the expectations are calculated using the discrete copula density $D_1 A D_2$. Define

$$L_l(\lambda_1, \dots, \lambda_k) := \sum_{i=1}^n \sum_{j=1}^n d^{(1)}(u_i) d^{(2)}(v_j) A(u_i, v_j) h_l(u_i, v_j) - \alpha_l, \quad l = 1, 2, \dots, k. \quad (5)$$

We seek the simultaneous roots of these functions and so minimize

$$L_{sum}(\lambda_1, \dots, \lambda_k) = \sum_{l=1}^k L_l^2(\lambda_1, \dots, \lambda_k).$$

The problem can be solved using one of Matlab's optimization procedures, FMINSEARCH, which implements the Nelder-Mead simplex method.⁽⁴²⁾ This is used in the examples in this article.

We remark that, given the choice of functions (h_1, \dots, h_k) , we have a parametric class of distributions with parameters the expected values $(\alpha_1, \dots, \alpha_k)$ of (h_1, \dots, h_k) . However, although we have a parametric family, we do not have a closed-

form expression for that family. Although the kernel in Equation (4) has a closed-form expression, the functions $d^{(1)}$ and $d^{(2)}$ do not. They are, however, uniquely defined and simple to compute. Pseudocode is given in the Supporting Information to the article.

When fitting common parametric copulas such as the t -copula using expert judgment it can be difficult to relate the parameters of the copula to observable quantities for which we can ask experts values. This is not true using minimum information copulas, however, due to the flexibility of the functions $h_i(\cdot)$. As an example, we show how an expert could specify a copula though defining two expected values.

3.3. Example

Suppose X and Y represent the failure times of two components that are functionally identical and physically colocated. There are many reasons to believe that the distributions of X and Y will be dependent, but modeling all the different sources of dependency may be difficult to do explicitly. Assume that the marginal distribution functions F_X and F_Y are exponential with mean time to failure 100, and that we want to specify a copula for (X, Y) .

We could ask an expert for information about the likelihood of near-simultaneous failure. Suppose that the expert assesses that the probability of both systems failing within time 1 of each other is 0.1, and that the probability both systems fail within time 10 of each other is 0.3. The expert information says that if we consider the functions of the copula variables U and V , defined by:

$$h_1(u, v) = \begin{cases} 1 & \text{if } |F_X^{-1}(u) - F_Y^{-1}(v)| < 1 \\ 0 & \text{otherwise,} \end{cases}$$

$$h_2(u, v) = \begin{cases} 1 & \text{if } |F_X^{-1}(u) - F_Y^{-1}(v)| < 10 \\ 0 & \text{otherwise,} \end{cases}$$

then the copula needs to be chosen so that the expected value of h_1 is 0.1 and that of h_2 is 0.3. Using the methods discussed here we can construct the minimum information copula.

In general, the range of expectation values available to the expert will be constrained, in the first instance by the choice of marginal distribution, and then by the expected value chosen for the first function. This was discussed by Bedford⁽²⁸⁾ in the context of a single copula. Two important aspects are discussed there. First, we can choose functions for evaluation that have a real "operational meaning"

for the experts, which is better than asking them to assess moments or abstract parameter values. Second, as the range of possible values for the expectation of a function can be computed by evaluating the function's expected value as we change the Lagrange multiplier values in Expression (4), we can offer guidance to experts about what values may be chosen to be consistent with those already chosen.

The resulting parameter values for this copula, found using a discretized grid of 200×200 points, are $\lambda_1 = -12.9100$ and $\lambda_2 = -1.377$. The left-hand side of Fig. 3 shows the minimally informative copula for these values.

On the right-hand side of Fig. 3, we have included a contour plot of the copula density transformed to allow for standard normal margins by transforming the copula coordinates (u, v) to (z_1, z_2) with $z_j = \Phi^{-1}(u_j)$ for $j = 1, 2$. This allows us to assess which is the closest bivariate copula to the minimum information copula fitted and so allows comparison to common parametric copulas. In the case of a Gaussian copula, the contour plots will be elliptical, while shapes like pears give indication of tail dependence induced, for example, by a Clayton or Gumbel copula. Bivariate t -copulas are identified through diamond-shaped contours. In this case, we see an elliptical shape.

4. COPULA COMPACTNESS

The previous section showed how bivariate minimum information copulas can be constructed and provides a useful family of bivariate copulas. However, the article aims to construct higher dimensional copulas. An important technical step is taken in this section where we consider the amount of variability between different bivariate copulas arising in a multivariate copula. The key step is to show that the family of bivariate (conditional) copula densities contained in a given multivariate copula distribution forms a compact set in the space of continuous functions on $[0, 1]^2$. We can then show that the same finite parameter family of copulas can be used to give a given level of approximation to all conditional copulas simultaneously.

It is important to define precisely the way in which we approximate densities. We assume that all densities are continuous and uniformly bounded away from zero. Write $C(Z)$ for the space of continuous real-valued functions on a space Z , where we shall always take $Z = [0, 1]^r$ for some r . A norm on the space $C(Z)$ is given by:

$$\|f_{1\dots r}\| = \sup |f_{1\dots r}(x_1, \dots, x_r)|, \quad f_{1\dots r} \in C(Z).$$

Since our functions are assumed continuous on Z , and since Z is compact, the norm of any such function is finite. We shall be interested in the set of all possible two-dimensional (conditional) copulas associated to a given continuous density function f :

$$\mathcal{C}(f) = \{c_{ij|i_1\dots i_r} : 1 \leq i, j, i_1, \dots, i_r \leq n, i, j \neq i_1, \dots, i_r\},$$

where $c_{ij|i_1\dots i_r}$ is the copula of the conditional density of X_i, X_j given X_{i_1}, \dots, X_{i_r} . Thus, $\mathcal{C}(\cdot)$ is an infinite set. It will be important to show that this set is relatively compact in the space of all continuous real-valued functions $C([0, 1]^2)$ because then we can show that the copula densities can be uniformly approximated. We consider compactness relative to the topology induced by the sup norm.

Compactness of a set K can be defined equivalently through one of two properties, each of which we shall use. (1) Any open cover of K has a finite subcover. In other words, if K is a subset of an infinite union of open sets, then it is in fact also a subset of a finite union of those open sets. (2) Any sequence of points (which in our case are functions) of K has a convergent subsequence.

The Arzela-Ascoli Theorem gives another way of checking compactness when dealing with function spaces. It says that a subset $K \subset C([0, 1]^2)$ is relatively compact if the functions of K are equicontinuous and point-wise bounded. We recall that a set of functions is equicontinuous if for all $\epsilon > 0$ and (u, v) there is a $\delta > 0$ such that if the Euclidean distance $|(u, v) - (u', v')| < \delta$ then $|g(u, v) - g(u', v')| < \epsilon \forall g \in K$, and that K is point-wise bounded if $\sup\{\|g\| : g \in K\} < \infty$.

As a first step to showing the relative compactness of $\mathcal{C}(f)$, we first consider two other spaces: the set of conditional marginal densities:

$$\mathcal{M}(f) = \{f_{i|i_1\dots i_r} : 1 \leq i, i_1, \dots, i_r \leq n, i \neq i_1, \dots, i_r\},$$

where $f_{i|i_1\dots i_r}(x_i | x_{i_1}, \dots, x_{i_r}) : [0, 1] \rightarrow \mathbb{R}$ are the conditional densities of X_i given X_{i_1}, \dots, X_{i_r} , one function for each combination of conditioning values x_{i_1}, \dots, x_{i_r} , and the set of conditional bivariate densities:

$$\mathcal{B}(f) = \{f_{ij|i_1\dots i_r} : 1 \leq i, j, i_1, \dots, i_r \leq n, i, j \neq i_1, \dots, i_r\},$$

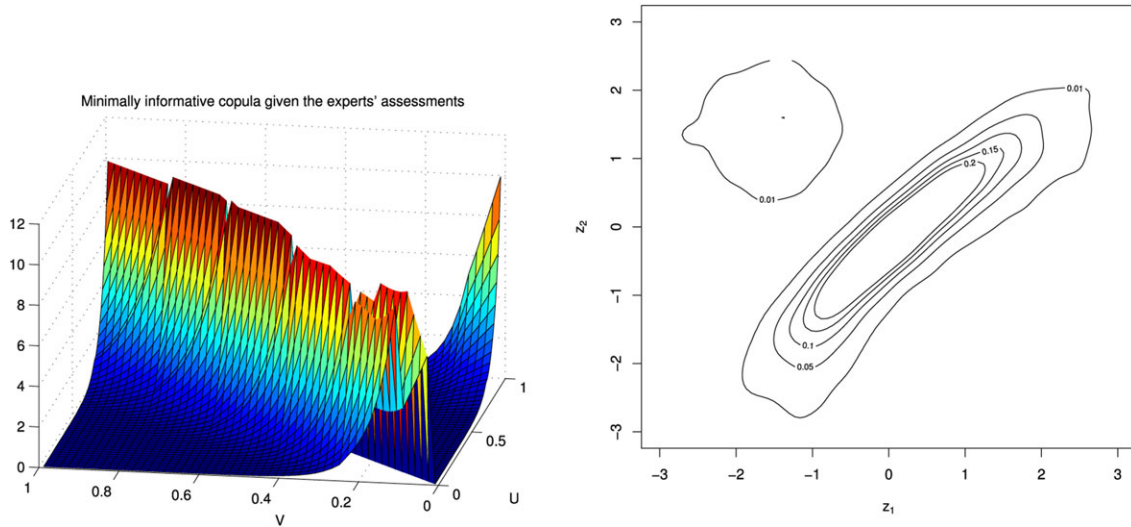


Fig. 3. A plot of the minimum information copula and transformed contour plot for X, Y .

where $f_{ij|i_1 \dots i_r}$ is the conditional density of X_i, X_j given X_{i_1}, \dots, X_{i_r} . Thus, $\mathcal{M}(f), \mathcal{B}(f)$ are also infinite sets. As we have defined it, a member of $\mathcal{M}(f)$ is a function of one variable—in other words, all the different marginals that we get for different conditions are individually members of $\mathcal{M}(f)$. Similarly for $\mathcal{B}(f)$. Hence, $\mathcal{M}(f) \subset C([0, 1])$ and $\mathcal{B}(f) \subset C([0, 1]^2)$.

THEOREM 2. *The sets $\mathcal{M}(f) \subset C([0, 1])$ and $\mathcal{B}(f) \subset C([0, 1]^2)$ are relatively compact.*

Proof. See Appendix A. \square

THEOREM 3. *The set $\mathcal{C}(f) \subset C([0, 1]^2)$ is relatively compact.*

Proof. See Appendix A. \square

Since all the functions in $\mathcal{C}(f)$ are positive and uniformly bounded away from 0 it follows that:

COROLLARY 1. *The set $\mathcal{LNC}(f) = \{\ln(g) : g \in \mathcal{C}(f)\} \subset C([0, 1]^2)$ is relatively compact.*

4.1. Linear Bases and Approximate Copulas

Consider the ordered set of sequences $h_0, h_1, h_2, \dots \subset C([0, 1]^2)$. We would like any finite sequence h_0, h_1, \dots, h_n to be linearly independent modulo the constants. The set $C([0, 1]^2)$ can be considered a vector space. Define V_n to be the vector space generated by the first n terms in the sequence. We would also like to show that $\cup_n V_n$ is dense in $C([0, 1]^2)$.

A countable basis h_0, h_1, \dots of $C([0, 1]^2)$ over the field \mathbb{R} is a countable subset $h_0, h_1, \dots \subset C([0, 1]^2)$ with the property that every element $v \in C([0, 1]^2)$ can be written as an infinite series $v = \sum_{i=0}^{\infty} \lambda_i h_i$, in exactly one way, where $\lambda_i \in \mathbb{R}$.

Consider the countable basis h_0, h_1, \dots . Since $v = 0$ can be written in exactly one way, then this must be with $\lambda_i = 0$ for all i . This means that any finite collection of basis elements is linearly independent modulo the constants. If we set $h_0 = 1$, then, for any n , h_1, \dots, h_n are linearly independent. It is also clear that $\cup_n V_n$ is dense in $C([0, 1]^2)$. There are lots of possible bases, for example, $u, v, uv, u^2, v^2, u^2v, uv^2, \dots$.

Given an ordered basis $h_1, h_2, \dots \in C([0, 1]^2)$ and a required degree of approximation $\epsilon > 0$ in the sup metric, we can consider the collection of open sets:

$$U_{k,\epsilon} = \{g \in C([0, 1]^2) : \inf \|g - \sum_{i=1}^k \lambda_i h_i\| < \epsilon\},$$

where the infimum in the above definition is to be taken over all possible values of the λ_i . Now, $U_{k,\epsilon}$ is clearly open and furthermore:

$$U_{k,\epsilon} \subset U_{k+1,\epsilon}, \quad \bigcup_{k=1}^{\infty} U_{k,\epsilon} = C([0, 1]^2).$$

So the $U_{k,\epsilon}$ form an open cover of $\mathcal{LNC}(f)$ and hence by definition of compactness there is a k such that $U_{k,\epsilon}$ covers $\mathcal{LNC}(f)$. We can state this as a result.

THEOREM 4. *Given $\epsilon > 0$, there is a k such that any member of $\mathcal{LNC}(f)$ can be approximated to within error $\epsilon > 0$ by a linear combination of h_1, h_2, \dots, h_k .*

The same result holds for $\mathcal{C}(f)$ (though not necessarily with the same k). We call the linear combination $\sum_{i=1}^k \lambda_i h_i$ an approximate copula because it is not guaranteed to be a copula itself. The next section shows that it can be adjusted slightly to obtain a copula that provides good approximation.

We remark that though we have been looking at approximation in the sense of the sup norm, one could easily look at higher order approximation. For example, if we assume that the density $f_{1..n}$ is continuously differentiable, then all the derivatives are continuous functions and the same arguments as used above show that they form an equicontinuous and point-wise bounded family. Following through we find that the copulas generated from $f_{1..n}$ are also continuously differentiable. By using a slightly different norm on the continuously differentiable functions $C^1([0, 1]^2) \subset C([0, 1]^2)$,

$$\|g\|_1 = \|g\| + \left\| \frac{d}{du} g \right\| + \left\| \frac{d}{dv} g \right\|,$$

we can guarantee that a similar approximation result to the above holds with point-wise approximation of the derivatives.

4.2. Ensuring that Approximating Densities are Copula Densities

Since the approximation we make of a copula density is not guaranteed to be a copula density itself, we need to transform it to obtain a copula. This is done by weighting the density as described in Section 3.2. If we have a continuous positive real-valued function $A(u, v)$ on $[0, 1]^2$, then there are continuous positive functions $d^{(1)}(u)$ and $d^{(2)}(v)$ such that $d^{(1)}(u)d^{(2)}(v)A(u, v)$ is a copula density, that is, it has uniform marginals. We call this density the *C-Projection* of A and denote it $C(A)$. It will be convenient to denote by $N(h)$ the normalization of a non-negative function h with finite integral.

We can control the error made when approximating a copula by another function.

Lemma 1. *Let g be a nonnegative continuous copula density. Given $\epsilon > 0$ there is a δ such that if $\|g - f\| < \delta$ then $\|g - C(f)\| < \epsilon$.*

Proof. See Appendix A. \square

The reweighting functions have the same differentiability properties as the function f being reweighted. This can be seen from the integral equation that they satisfy:

$$d^{(1)}(u) = \frac{1}{\int d^{(2)}(v) f(u, v) dv} \quad \text{and} \\ d^{(2)}(v) = \frac{1}{\int d^{(1)}(u) f(u, v) du}.$$

We use Equation (2) to see that good approximation of each conditional copula gives a good approximation of the multivariate density.

5. CONSTRUCTING APPROXIMATIONS USING MINIMALLY INFORMATIVE DISTRIBUTIONS

The above discussion has shown that we can approximate all conditional copulas using linear combinations of basis functions. We did not address the question of how you choose the appropriate parameter values, and finding the parameters that would minimize the sup norm for a given copula is not an appealing procedure. An alternative that lies very close to the approach described above is to use the minimum information criterion. Given $\{1, h_1, \dots, h_k\} : [0, 1]^2 \rightarrow \mathbb{R}$ we seek values $\lambda_1, \dots, \lambda_k$ so that $\exp(\sum_{i=1}^k \lambda_i h_i)$ is close to the copula density we are approximating.

In the minimum information framework, we do this by fitting the moments of h_i . So if $\int \int h_i g du dv = \alpha_i$ then we search for the copula density with minimum information (with respect to the independent distribution) that has those moments. This copula density is unique and has the form:

$$d^{(1)}(u)d^{(2)}(v) \exp\left(\sum_{i=1}^k \lambda_i h_i(u, v)\right).$$

When we use a vine structure to model a multivariate distribution, the vine defines a decomposition of the multivariate distribution into conditional copulas, associated to the conditioned and conditioning sets of the vine. For example, if $\{i, j\}$ is the conditioned set and D_e is the conditioning set in one part of a vine, then the family of conditional copulas for x_i, x_j given D_e has to be specified. Using the minimum information approach means that we specify mean values for the functions h_r given the variables in D_e , that is, we specify the conditional means $\alpha_m(ij \mid D_e)$.

A multivariate distribution can be approximated as follows. Specify a basis family $\mathcal{B}(k)$, specify a vine structure and for each part of vine, specify either expected values $\alpha_1, \dots, \alpha_k$ for h_1, \dots, h_k on each pairwise copula or functions $\alpha_m(ji \mid D_e)$ for the expected values as functions of the conditioning variables, for $m = 1, \dots, k$.

We remark that, since under our assumptions there is a uniform lower bound on the density of the copulas used in the representation, the uniform point-wise approximation that can be achieved implies information convergence in two ways. By making the copula approximations close enough, we can ensure that (i) the information of the overall approximate multivariate copula (with respect to the independent copula) is close to that of the original multivariate copula, and (ii) the information of our approximate multivariate copula relative to the original multivariate copula is close to zero.

We illustrate the procedure by applying it to two financial data sets.

5.1. Example: Stock Market Time Series

We use the same data set as considered by Aas *et al.*⁽¹¹⁾ We have four time series of daily data: the Norwegian stock index (TOTX), the MSCI world stock index, the Norwegian bond index (BRIX), and the SSBWG hedged bond index. All are for the period January 1, 1999 to July 8, 2003. We denote these four variables T , B , M , and S .

We generate a vine approximation fitted to this data set using minimum information distributions. We adopt a vine structure similar to that in Fig. 1 with variables T, B, M, S being 1, 2, 3, 4, respectively. We can find the corresponding functions of the copula variables X, Y, Z , and W associated with T, M, B, S . These are defined by, for example, $h_i(X, Y) = h'_i(F_1^{-1}(T), F_2^{-1}(M))$ and have the same specified expectation, in this case $E[h'_i(T, M)] = E[h_i(X, Y)]$. The minimum information copulas calculated are derived based on the copula variables X, Y, Z, W .

Initially, we construct minimally informative copulas between each set of two adjacent variables in the first tree, T_1 . We must decide on which bases to take and how many discretization points to use in each case. We illustrate the recommended procedure for the first copula in T_1 , between T, M .

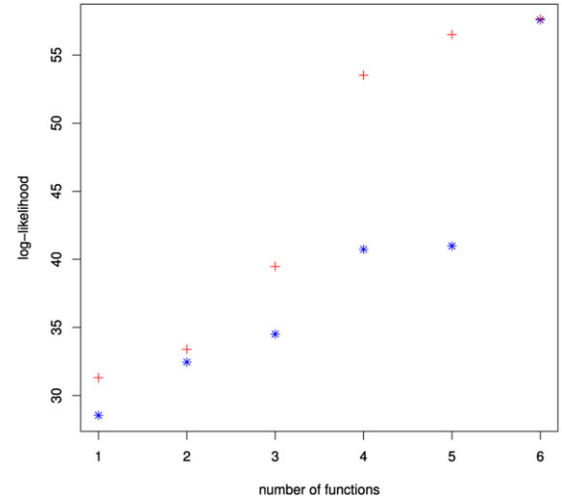


Fig. 4. The log-likelihood of the minimally informative copula calculated based on different functions for the simple (blue stars) and stepwise (red crosses) methods (colors visible in on-line version).

5.1.1. Step-wise Inclusion of Basis Functions

We wish to know which basis functions to include in our copula. We could choose basis functions, starting with simple polynomials and moving to more complex ones, and include them until we are satisfied with our approximation. For example, if we included the following basis functions in order $TM, TM^2, T^2M, TM^3, T^3M, T^2M^3$, then the log-likelihood for the copula changes as in the blue stars in Fig. 4.

There is a jump in the log-likelihood as we add the sixth basis function, T^2M^3 . This could imply that we are not adding the basis functions in an optimal manner. Instead, at each stage, we propose to assess the log-likelihood of adding each additional basis function. We include the function that produces the largest increase in the log-likelihood. Our method is similar to a step-wise regression. Doing so for the initial copula yields the basis functions $TM^2, T^2M, T^2M^2, TM, TM^4, T^2M^4$. The log-likelihood at each stage is given in the red crosses in Fig. 4.

We see that there is no longer a jump in the log-likelihood when adding the sixth basis function. The log-likelihood also increases more quickly and reaches its plateau value of around 60 using fewer basis functions. We use this step-wise technique to choose all of the remaining basis functions in the example. The use of log-likelihood in this way is not inconsistent with minimum information modeling.

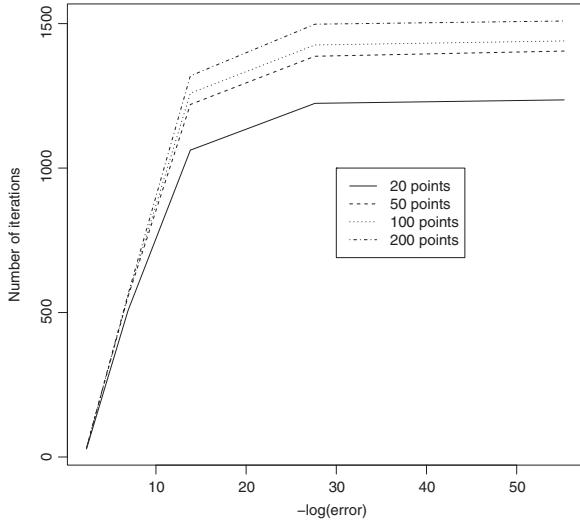


Fig. 5. A plot of the number of iterations against convergence level for 20, 50, 100, and 200 discretization points.

Jaynes⁽²⁵⁾ uses the parameter maximum likelihood estimates associated with the form of the minimum information distribution to justify the connection in the constraint rule of expectations and frequencies.

5.1.2. Returning to the Example

We include the six basis functions given above, that is, $h'_1(T, M) = TM^2$, $h'_2(T, M) = T^{2M}$, $h'_3(T, M) = T^2M^2$, $h'_4(T, M) = TM$, $h'_5(T, M) = TM^4$, and $h'_6(T, M) = T^2M^4$. We shall fix the values of the expectations of these functions by using the empirical data. For example, $\alpha_1 = \frac{1}{1,094} \sum_{i=1}^{1,094} t_i m_i^2$, as there are 1,094 observations for each variable.

The minimum information copula C_{TM} with respect to the uniform distribution given the six constraints above can be constructed. To do so, we need to decide on the number of discretization points (or grid size). A larger grid size will provide a better approximation to the continuous copula but at the cost of more computation time. Similarly, the more iterations of the D_1AD_2 and optimization algorithms that are run, the more accurate the approximation will become. This is again at the expense of speed. Comments on the D_1AD_2 algorithm are given in Section 3.2. In terms of the optimization, we can specify how accurate we wish our approximation to be and then judge the effect on the number of iterations required for convergence. The number of iterations needed will also depend on the grid size.

Fig. 5 gives a plot comparing the number of iterations required for convergence of FMINSEARCH

Table I. Constraints and Parameter Values for C_{MB} and C_{BS}

$h'_i(M, B)$	α_i	λ_i	$h'_i(B, S)$	α_i	λ_i
MB	0.2905	24.970	BS	0.2375	18.818
M^2B	0.2066	-22.233	B^2S	0.1546	-26.914
M^3B	0.1611	20.308	B^3S	0.1142	7.929
M^2B^3	0.1223	32.006	B^3S^2	0.0730	-13.949
M^2B^2	0.1527	-39.639	BS^2	0.1537	-24.939
MB^5	0.1142	-3.910	B^2S^2	0.0992	36.763

given a certain error of L_{sum} and grid size. The errors considered are in the range 1×10^{-1} to 1×10^{-24} . These are then transformed by taking $-\log(\cdot)$ and this is the quantity plotted.

We see that the larger the number of grid points used, the larger the number of iterations needed for convergence. This is true over all error levels. The grid sizes all follow the same pattern, with large increases in the number of iterations needed for improved accuracy initially and smaller increases when the error is smaller.

Throughout the rest of the example, we choose a grid size of 200×200 and shall work to an error of 1×10^{-12} . This corresponds to a transformed error in Fig. 5 of 27.63. This represents a suitable balance between providing an accurate approximation to the minimally informative copula and keeping computational effort to a reasonable level.

We can find the minimally informative copula C_{TM} . Pseudo-code for doing this is given in the Supporting Information. This gives parameter values of $\lambda_1 = 17.0262$, $\lambda_2 = -17.6367$, $\lambda_3 = -1.1117$, $\lambda_4 = 4.7746$, $\lambda_5 = -26.8054$, $\lambda_6 = 19.9014$. The copula is plotted on the left-hand side of Fig. 6 and the contour plot of the copula density transformed to allow for standard normal margins is given on the right-hand side. The log-likelihood for the copula is $l_{TM} = 58.1256$.

The remaining copulas in T_1 are C_{MB} , C_{BS} . The constraint functions, constraints, and Lagrange multipliers used for this copula are given in Table I. The log-likelihoods are $l_{MB} = 155.18$ and $l_{BS} = 19.23$, respectively.

The conditional copulas in the second tree, T_2 , can be approximated using the minimum information approach. Initially, we construct the conditional minimum information copula between $T | M$ and $B | M$. Aas *et al.*⁽¹¹⁾ considered the dependencies between T and B given M to be constant over $M \in [0, 1]$.

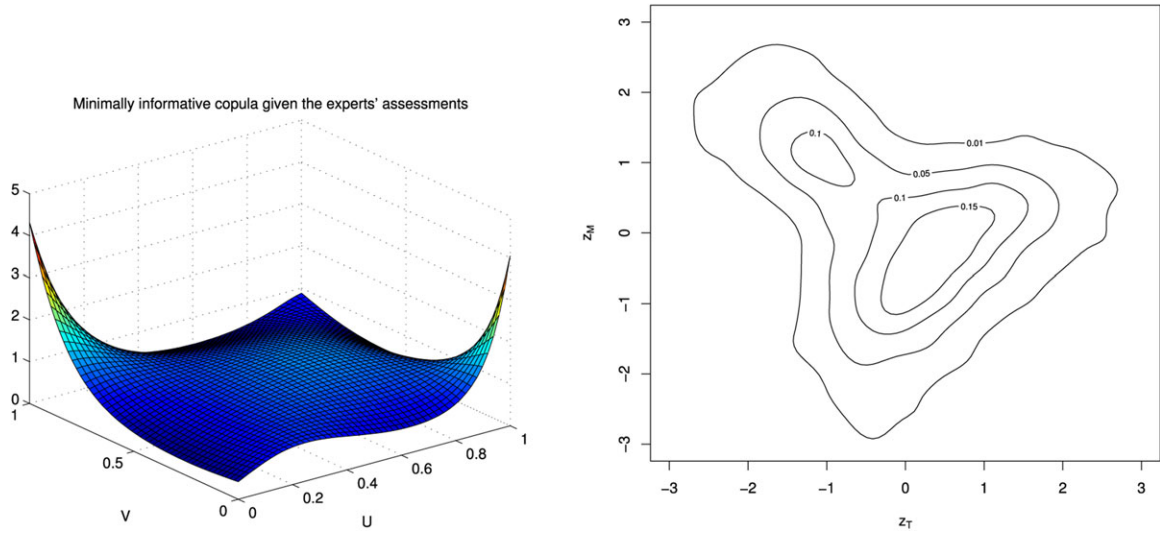


Fig. 6. The minimally informative copula between T and M and transformed contour plot, Norwegian stock data.

Instead, we divide the support of M into some arbitrary subintervals or bins and then construct the conditional copula within each bin. We will investigate the effect of this in the following example. We find bases in the same way as for the marginal copulas and fit the copulas to the expectations calculated for these. We use four bins so that the first copula is for $T, B \mid M \in (0, 0.25)$. The bases for this copula are $h'_1(T, B \mid M \in (0, 0.25)) = T^2 B$, $h'_2(T, B \mid M \in (0, 0.25)) = T^3 B$, $h'_3(T, B \mid M \in (0, 0.25)) = T^4 B$, $h'_4(T, B \mid M \in (0, 0.25)) = T^5 B$, $h'_5(T, B \mid M \in (0, 0.25)) = T B^5$, $h'_6(T, B \mid M \in (0, 0.25)) = T^2 B^4$. The expectations given these basis functions that will constrain the minimum information copula are $\alpha_1 = 0.1246$, $\alpha_2 = 0.0983$, $\alpha_3 = 0.0813$, $\alpha_4 = 0.0693$, $\alpha_5 = 0.0239$, and $\alpha_6 = 0.0220$.

We follow this process again for the remaining bins. Table II shows the constraints and corresponding Lagrange multipliers required to build the conditional minimum information copula between $T \mid M \in (0, 1)$ and $B \mid M \in (0, 1)$. The overall log-likelihood of the conditional minimum information copula between T and B given $M \in (0, 1)$ is 29.242.

Similarly, we can construct the minimum information copula between $M \mid B$ and $S \mid B$ based on four bins and six constraints. The resulting minimum information copula has a log-likelihood of 16.3901.

The conditionally minimally informative copula in the third tree, T_3 , can be obtained. We first divide each of the conditioning variables' supports into four bins as in T_2 . Then, the minimum information copulas for $T \mid (M, B)$ and $S \mid (M, B)$ are calculated on each

combination of bins for M, B . In T_3 , there are 16 bins altogether. Details are omitted. The log-likelihood of T_3 is 110.69.

The log-likelihood of the overall vine, obtained by summing the log-likelihoods of each of the component copulas, is 388.859. This is larger than that using the vine construction of bivariate t -copulas and constant conditional dependence of Aas *et al.*⁽¹¹⁾ of 291.801. Suppose, rather than choosing our bases using the step-wise method, we had calculated all of the copulas using the same six basis functions. Further suppose that those chosen were the simple polynomials, XY , XY^2 , X^2Y , XY^3 , X^3Y , X^2Y^3 . Then the overall log-likelihood is 370.147. This is lower than when using our approach but better than the t -copula of Aas *et al.* However, the advantage of the step-wise method can be seen if we take fewer basis functions for each copula. If we take 5, we obtain a log-likelihood of 377.552, which is still larger than that obtained using 6 without the step-wise approach.

5.2. Example: Comparison with the Gaussian Copula

We consider five years of exchange rates against the U.S. dollar for four different currencies: the Great British Pound, the Euro, the Japanese Yen, and the South Korean Won. Before fitting the copula models to the data, we first remove any trends, seasonality, etc., from the data by fitting $ARMA(p, q) - GARCH(r, s)$ models to each of the individual time series. The analysis is then conducted on the

Table II. Bases, Parameter Values, and Log-Likelihoods for $C_{TB|M}$

Interval	Bases	Parameter Values
$0 < M < 0.25$	$(T^2B, T^3B, T^4B, T^5B, TB^5, T^2B^4)$	$(26.0, -141.5, 231.8, -120.0, 12.4, 10.6)$
$0.25 < M < 0.5$	$(TB, TB^2, T^3B, T^4B, T^2B, T^2B^3)$	$(-32.4, 16.0, -188.2, 112.2, 103.3, -9.2)$
$0.5 < M < 0.75$	$(T^2B, TB^2, T^3B, T^2B^3, TB, TB^3)$	$(13.4, 33.6, 12.1, -22.2, -35.0, -4.2)$
$0.75 < M < 1$	$(TB^2, TB^3, TB^4, TB, T^5B, TB^2)$	$(-22.5, 38.5, -23.6, 1.7, -3.6, 6.7)$

empirical cdf values of the residuals from the time-series models. For more details on this, see Ref. 11.

In order to fit a four-dimensional D-vine to the data, we need to identify a structure for the vine. Using the methods in this article, we know that we can fit any vine structure arbitrarily well using bivariate minimum information copulas. However, we select the structure using the method given in the VineCopula package in R. This identifies the structure of the vine sequentially, modeling the strongest correlations in the first tree of the vine, assuming that the bivariate copulas do not change with the conditioning value. Further information is given in Ref. 19.

The resulting structure of the D-vine gives Euros, Great British Pounds, South Korean Won, and Japanese Yen, respectively, in the first tree. We relabel these currencies 1, 2, 3, and 4.

The Gaussian copula has been criticized for its widespread use in the financial sector in spite of evidence that the assumptions underlying modeling and necessary for use were not being met.⁽⁴³⁾ One such assumption of the Gaussian copula is that the conditional dependencies between variables in the model are constant. We apply the Gaussian copula, as well as a minimum information vine structure, to the exchange rate data to investigate the suitability of this assumption.

The four-dimensional Gaussian copula for the currencies takes the form:

$$C_{\Sigma}(x_1, x_2, x_3, x_4) = \Phi_{\Sigma}(\Phi^{-1}(F_1(x_1)), \Phi^{-1}(F_2(x_2)), \Phi^{-1}(F_3(x_3)), \Phi^{-1}(F_4(x_4))),$$

where $\Phi_{\Sigma}(\cdot, \cdot, \cdot, \cdot)$ is the cdf of the 4-variate standard Gaussian distribution with mean zero and variance matrix Σ , $\Phi^{-1}(\cdot)$ represents the inverse cdf of the univariate standard Gaussian distribution, and $F_i(\cdot)$ represents the cdf for currency $i = 1, 2, 3, 4$.

We first fit a Gaussian copula to the residual series. The fitted values for the correlations are:

$$\begin{aligned} \rho_{12} &= 0.61, \rho_{13} = 0.31, \rho_{14} = -0.058, \rho_{23} = 0.35, \\ \rho_{24} &= 0.027, \rho_{34} = -0.081. \end{aligned} \quad (6)$$

Table III. The Constraints and Lagrange Multipliers for the Three Marginal Copulas in the First Tree of the Vine

Copula Variables	$(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$	$(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$
u_1, u_2	$(0.301, 0.218, 0.219, 0.166)$	$(33.63, -20.15, -33.90, 30.17)$
u_2, u_3	$(0.280, 0.196, 0.197, 0.142)$	$(26.22, -21.69, -22.49, 22.21)$
u_3, u_4	$(0.244, 0.162, 0.159, 0.105)$	$(21.36, -25.22, -18.89, 21.88)$

We fit a minimum information vine and compare the two approaches by simulating from the two distributions. We fit a D-vine in four dimensions. This requires a minimum information copula specified between exchange rates 1 and 2, one between 2 and 3, and another between rates 3 and 4, a conditional copula between rates 1 and 3 given exchange rate 2 and between rates 2 and 4 given exchange rate 3, and a conditional copula between rates 1 and 4 given exchange rates 2 and 3. We specify four basis functions for each copula and use the same basis functions each time, namely, $h_1(u_i, u_j) = u_i u_j$, $h_2(u_i, u_j) = u_i^2 u_j$, $h_3(u_i, u_j) = u_i u_j^2$, $h_4(u_i, u_j) = u_i^2 u_j^2$ for $i = 1, 2, 3, j \neq i$. We could have used the method from the previous example to choose the optimal basis functions.

Table III gives a summary of the constraints and resulting parameter values for the marginal copulas. The copulas are given in the top three plots of Fig. 7 and the contour plot of the copula density transformed to allow for standard normal margins are given, respectively, at the bottom of the figure.

We wish to split the support of u_2 into bins and define the conditional copula for u_1 and u_3 based on these bins. After plotting the conditional correlations for several different numbers of bins, we settle on four bins.

The remaining copula is that between u_1 and u_4 . To construct this, we must create bins of combinations of u_2 and u_3 . We separate u_2 and u_3 into four bins each, meaning that there are 16 combinations of these bins. We can calculate the correlations between

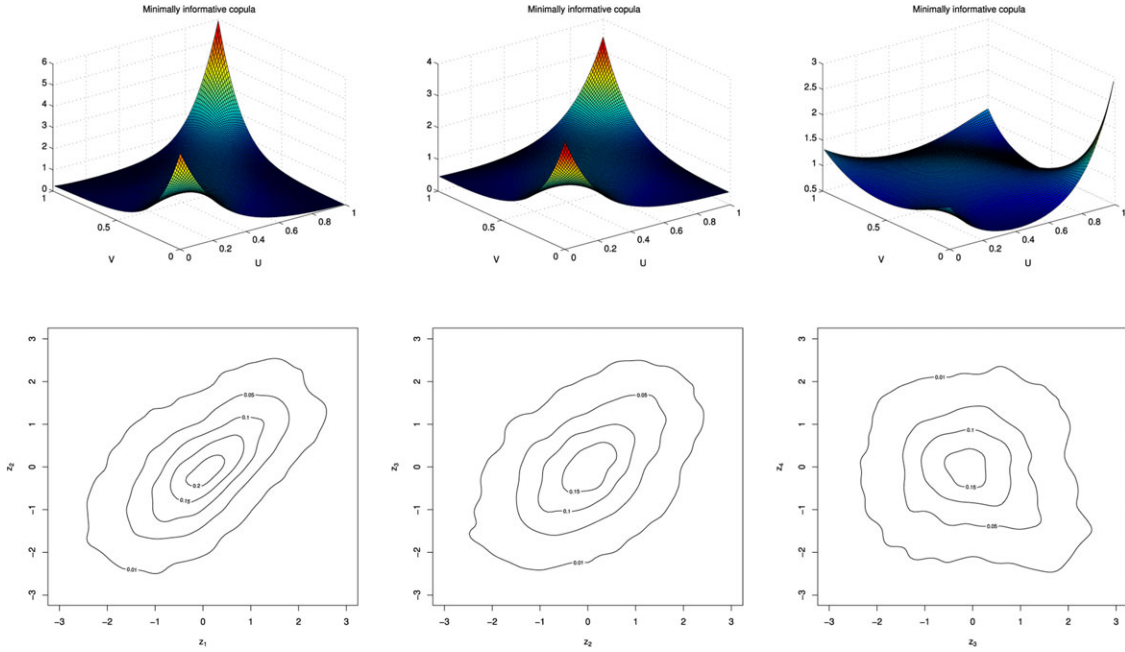


Fig. 7. The bivariate minimum information copulas (top) and transformed contour plots (bottom) for the exchange rates of currencies 1 and 2, 2 and 3, and 3 and 4, respectively.

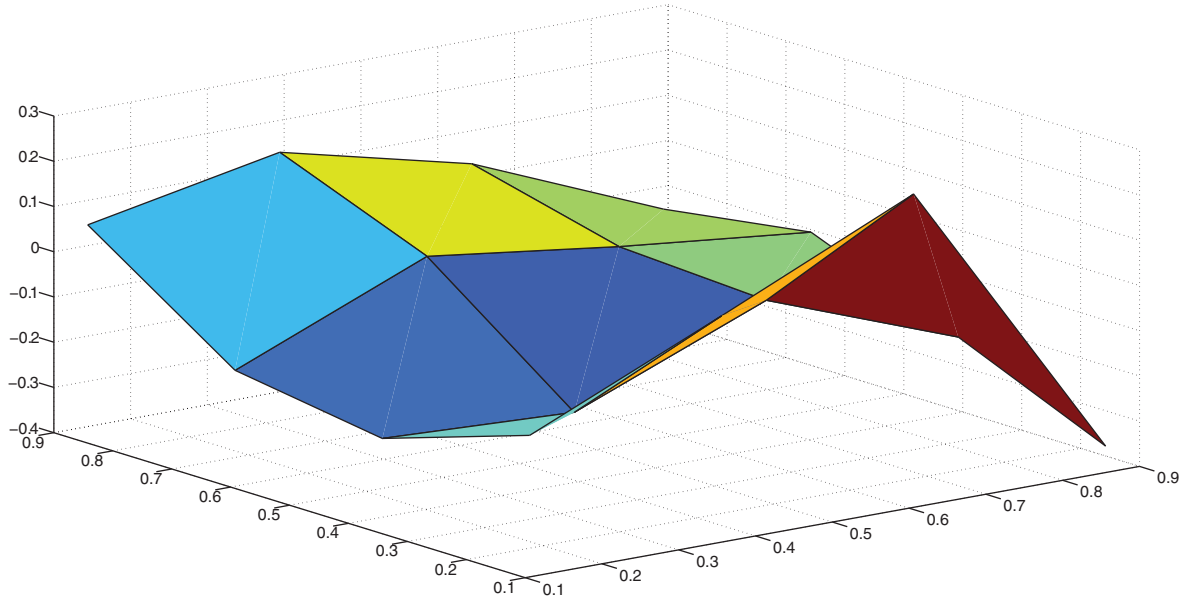


Fig. 8. The changes in conditional correlation between exchange rates 1 and 4 given different bins for exchange rates 2 and 3.

u_1 and u_4 for each of these bins and plot them as a surface, as in Fig. 8.

The empirical conditional expectation data are not inconsistent with the conditional correlation being a smooth function. The use of a smooth curve

to represent the conditional correlation is a possible way of compressing the data more compactly.

The fitted Lagrange multipliers for the two conditional copulas in tree 2, given the binning of u_2 and u_3 , respectively, are given in Table IV.

Table IV. The Lagrange Multipliers for the Two Conditional Copulas in the Second Tree of the Vine

Copula Variables	Bin	$(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$	$(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$
u_1, u_3	$u_2 \in (0, 0.25)$	(0.119, 0.071, 0.065, 0.041)	(48.03, -48.67, -51.07, 53.69)
u_1, u_3	$u_2 \in (0.25, 0.5)$	(0.188, 0.116, 0.100, 0.063)	(24.57, -21.39, -27.69, 25.28)
u_1, u_3	$u_2 \in (0.5, 0.75)$	(0.309, 0.215, 0.207, 0.145)	(12.98, -10.54, -9.12, 9.71)
u_1, u_3	$u_2 \in (0.75, 1)$	(0.488, 0.372, 0.400, 0.307)	(81.61, -76.40, -72.33, 69.16)
u_2, u_4	$u_3 \in (0, 0.25)$	(0.205, 0.146, 0.123, 0.089)	(49.86, -50.31, -51.80, 52.59)
u_2, u_4	$u_3 \in (0.25, 0.5)$	(0.216, 0.140, 0.126, 0.083)	(37.30, -35.93, -40.25, 39.04)
u_2, u_4	$u_3 \in (0.5, 0.75)$	(0.282, 0.185, 0.197, 0.131)	(59.97, -61.00, -57.31, 59.77)
u_2, u_4	$u_3 \in (0.75, 1)$	(0.307, 0.204, 0.233, 0.157)	(63.21, -64.21, -58.57, 60.01)

We fit the minimum information copulas for the different bins defined in the third tree, between u_1 and u_4 given u_2 and u_3 . The resulting Lagrange multipliers are found as previously. Details are omitted. This fully defines the vine.

5.2.1. Comparison of Models Using Simulation

To compare how well the two methods considered recover the structures within the data, we simulate from each model.

In the case of the minimum information vine, sampling from the constructed distribution can be carried out using the cumulative approach⁽¹⁴⁾ (also known as conditional sampling and the inverse of the Rosenblatt transform). The sampling strategy is as follows: sample four independent variables distributed uniformly on interval $[0, 1]$, denoted by W_1, W_2, W_3, W_4 , and calculate values of correlated variables X_1, X_2, X_3, X_4 by taking $x_1 = w_1$, $x_2 = F^{-1}(w_2 | x_1)$, $x_3 = F^{-1}(w_3 | x_1, x_2)$, and $x_4 = F^{-1}(w_4 | x_1, x_2, x_3)$, where x_i and w_i are realization values of X_i and W_i , respectively. Pseudocode for the sampling, including the binning, is given in the Supporting Information.

Initially, we consider the unconditional correlations in the data that were given in Equation (6). Taking 1,000 samples from the Gaussian copula results in estimates of these correlations being $\rho_{12} = 0.619$, $\rho_{13} = 0.341$, $\rho_{14} = -0.092$, $\rho_{23} = 0.368$, $\rho_{24} = 0.010$, and $\rho_{34} = -0.090$. A simulation of 1,000 samples from the minimum information vine gives estimates of $\rho_{12} = 0.611$, $\rho_{13} = 0.364$, $\rho_{14} = -0.161$, $\rho_{23} = 0.446$, $\rho_{24} = -0.098$, and $\rho_{34} = -0.118$.

Both methods reproduce the overall correlation structure fairly well. However, we saw from Fig. 8 that the conditional correlations between the exchange rates were nonconstant. To see whether our two models are capturing these dependencies, we investigate using cobweb plots.

Initially, we consider the data. On the left-hand side of Fig. 9 we give a cobweb plot of the four uniform variables for all of the observed time points. We see the overall shape of the distribution. On the right-hand side of the figure, we see the conditional relationship between u_1, u_4 conditional on $u_2 \in [0, 0.25)$, $u_3 \in [0, 0.25)$. This relationship is fairly strong. For more information on cobweb plots, see Ref. 44.

Fig. 10 gives the same two cobweb plots for the Gaussian copula, in the top row, and the minimum information vine, in the bottom row. Both methods are capturing the overall structure of the distribution well. However, when we condition on u_2, u_3 , the Gaussian copula fails to capture the conditional relationship between u_1, u_4 . The minimum information vine reproduces a conditional structure that is much closer to that found in the data.

5.3. Simulation Study: Tail Dependence

Our general results show that vines formed from minimum information copulas can represent any dependence structure. A much investigated case of dependence in financial and extreme weather risk modeling is tail dependence. In this section, we study tail behavior of the minimum information copula for data simulated from parametric copulas with various tail dependency, including heavy, symmetric, and nonsymmetric tails. Of course, in the real world we compare models to data and not models to models. Nevertheless, the simulation study may provide additional insight. Initially, we utilize scatter plots and Kendall process plots (K-plots),⁽⁴⁵⁾ which detect bivariate dependence using the ranks of the data.

A simulation study is carried out. We initially investigate data with nonsymmetric tail behavior. It is known that the Clayton and Gumbel copulas have asymmetric tails. The first column of Fig. 11 shows a scatter plot of a random sample taken from a bivariate Clayton copula with

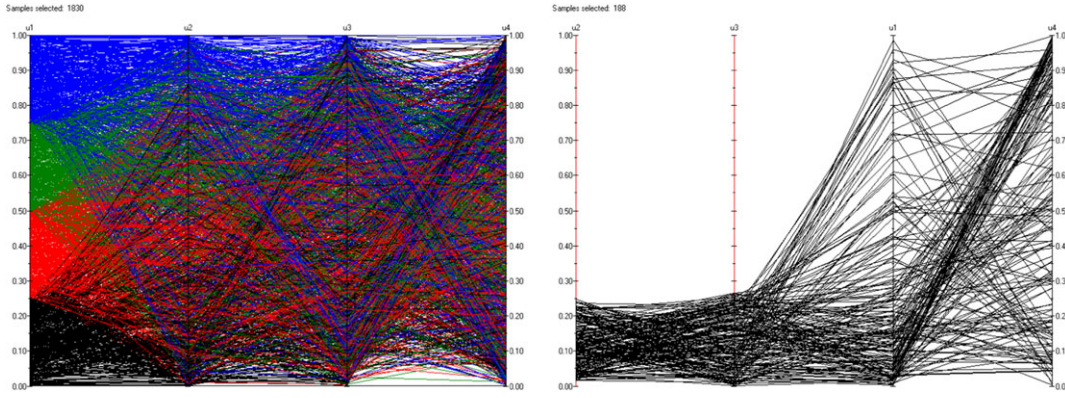


Fig. 9. Cobweb plots for all of the data values (left-hand side) and for u_1, u_4 conditional on $u_2 \in [0.75, 1)$, $u_3 \in [0, 0.25]$ (right-hand side).

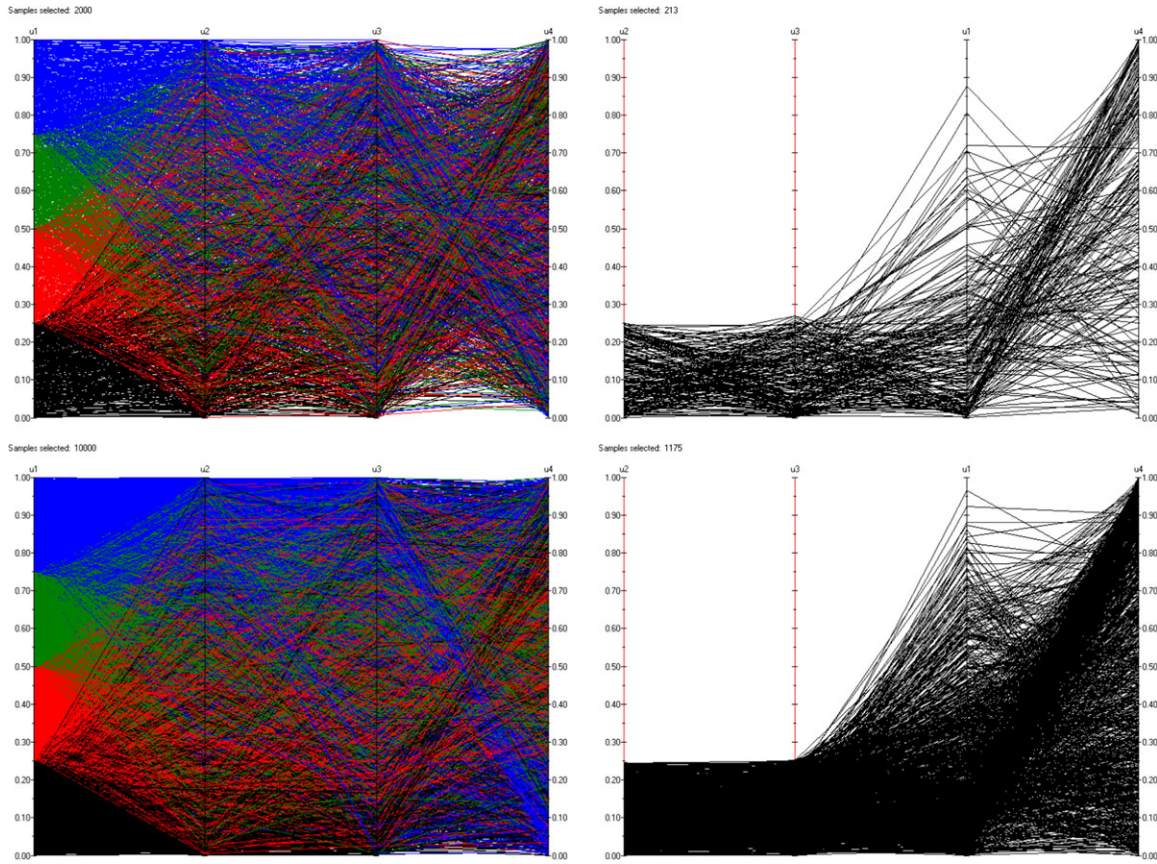


Fig. 10. Cobweb plots based on simulated data for the Gaussian copula (top row) and for the minimum information vine (bottom row).

parameter $\theta = 1.3979$, and the corresponding K-plot. The second column shows the same plots for a minimum information copula fitted to the sample drawn from the Clayton copula. The basis functions used are $U_1 U_2$, $U_1^2 U_2$, $U_1^3 U_2$, $U_1 U_2^2$, $U_1^4 U_2$ and

the resulting Lagrange multipliers are 9.02, 75.72, -63.81 , -20.98 , 7.54.

From the scatter plots, it appears that the minimum information copula is capturing the general behavior of the Clayton copula well. The upper

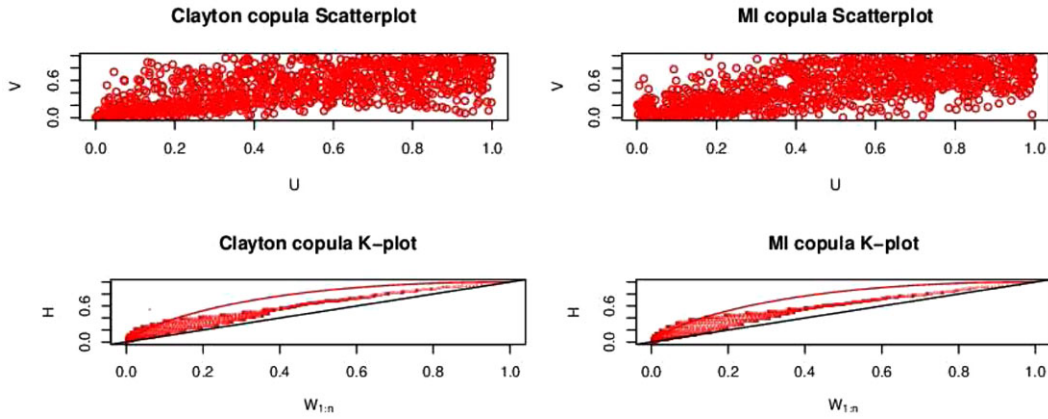


Fig. 11. Scatter- and K-plots of the Clayton copula and the fitted minimum information copula.

tail dependency behavior can be observed from the K-plots and in this case the Clayton and minimum information copula give similar plots. We have repeated the exercise for the Gumbel and t -copulas. The results are given in Appendix B. In all cases, the minimum information copula captures the data from the parametric copula well using the scatter plot and K-plot.

We extend our investigation to test the ability of minimum information bivariate copulas to capture the upper tail behavior found in t , Gumbel, and Tawn copulas. We will investigate the upper tail dependence coefficient. If our minimum information copula can successfully model the behavior in the upper-right corner of the unit square, then it will also be able to model the behavior in the lower left-hand corner.

We simulate pairs of 10,000 realizations from each of the parametric copulas identified above for each of 10 sets of parameter values. If θ is the first parameter of each copula and γ is the second parameter, then the values used for the t -copula are $\theta = (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$ (all with $\gamma = 4$ degrees of freedom) and for the Gumbel and Tawn copulas $\theta = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$, with $\gamma = 1$ always for the Tawn copula. The minimum information copula is fit to each of these pairs of observations using six polynomial basis functions of the type given in the examples in this article.

We use a nonparametric estimator^(46,47) of the upper tail coefficient for both the simulated data and the minimum information copula. If pairs of simulated values are (u_i, v_i) then this is:

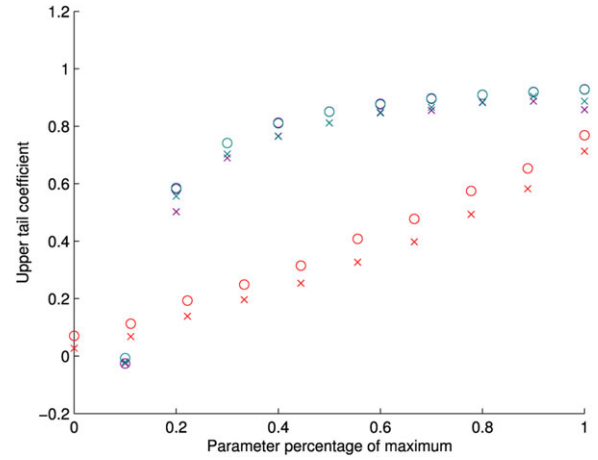


Fig. 12. Comparison of upper tail dependence coefficient for simulated values from the t (green), Gumbel (purple), and Tawn (gray) copulas and the minimum information copula. The circles represent the parametric copula and the crosses the minimum information copula in each case.

$$\hat{\lambda}_U = 2 - 2 \exp \left[\frac{1}{n} \sum_{i=1}^n \log \left\{ \sqrt{\log \left(\frac{1}{u_i} \right) \log \left(\frac{1}{v_i} \right)} / \log \left(\frac{1}{\max(u_i, v_i)^2} \right) \right\} \right].$$

We display the results of the simulation in Fig. 12. The different colors represent the simulated data from the different parametric families. The circles denote the upper tail coefficient for the simulated values and the crosses represent the coefficient for the minimum information copula fitted to the simulated data. In all cases, the estimated values for

the simulated data are close to the theoretical values for the parametric copula. On the x -axis, the parameter values from each of the copulas have been scaled to be a percentage of their maximum value in the simulation.

We see that the minimum information copula can model from weak to strong upper tail dependence in all of the parametric copulas chosen. This suggests it is suitable for modeling situations that display tail dependence. Given the basis functions used, however, the minimum information copula tends to underestimate the tail dependence coefficient slightly for all copulas over the ranges of parameter values. This could be improved by choosing basis functions that are concentrated in the top right-hand corner of the unit square. Future work will address this.

We consider tail-dependence in the multivariate case. To do so, we simulate tail dependent data in three dimensions (u_1, u_2, u_3) , fit a D-vine to the data with bivariate minimum information copulas, and investigate its ability to capture the conditional dependence between (u_1, u_3) given u_2 . We do not need to consider Gaussian or t -copulas as they are closed under bivariate marginalization. We consider the Gumbel copula. The empirical upper tail coefficients from varying parameter values and the estimated upper tail coefficients from the minimum information vine are given in Fig. 13. In each case 10,000 simulations were used.

We see a similar pattern to Fig. 12, with the minimum information approach able to capture high tail dependence between the variables but having a small downward bias in its estimates.

5.4. Discussion

5.4.1. Sources of Error

The method we have used has the theoretical property that it can be used to build arbitrarily good approximations to the original distribution. There are several sources of potential error in this approximation that we briefly discuss. One is the choice of base where it is convenient to take a low number of functions h_i . Using three terms $(u.v, u^2v, uv^2)$ rather than one (uv) , we are able to generate asymmetric copulas, which has value in modeling general data sets. The use of large numbers of functions does give more accuracy, at the cost of extra computation at the construction stage but at no extra cost at the sampling stage. The second general source of error arises from discretization. Discretization occurs dur-

ing the construction stage when we have to create data bins with which to calculate conditional expectations from the original data. There is a tradeoff, as with higher numbers of bins we have fewer data and noisier estimates of the conditional expectations. In our sampling, we simply used conditional expectations based on the bins, although we could have used them to estimate continuous functions. If we had used a joint normal distribution, then our implicit assumption would have been that these conditional expectations were all constant. We expect the error arising from this discretization to be the main source of errors overall. A final source of discretization error arises in the level of discretization in the D_1AD_2 algorithm. The algorithm works more slowly when using more points, but gives more accuracy. This becomes more important when the distribution being modeled is far away from uniform.

5.4.2. Vine Structure

We remark on the importance of vine structure. Any regular vine structure can be used to approximate any distribution. Many vine structures are available (see, for example, Kurowica and Cooke⁽¹⁴⁾), and even with a given vine structure we could permute the variables so that different conditional copulas were being considered in the construction. In view of the above discussion about error, we believe that it makes sense to consider different vine structures, and that a convenient measure of usefulness is the degree of complexity of the conditional expectation functions, which would ideally be constant or simple linear functions.

6. CONCLUSION

We present a novel method to approximate a multivariate distribution by any vine structure to any degree of approximation. We have operationalized the theoretical approximation results using minimum information copulas that can be specified to any required degree of precision based on the data available. We have shown rigorously that good approximation “locally” guarantees good approximation globally. This approximation allows the use of a fixed finite-dimensional family of copulas to be used in a vine construction, with the promise of a uniform level of approximation. That is, we can use the same bases to approximate each copula in each tree of the corresponding vine.

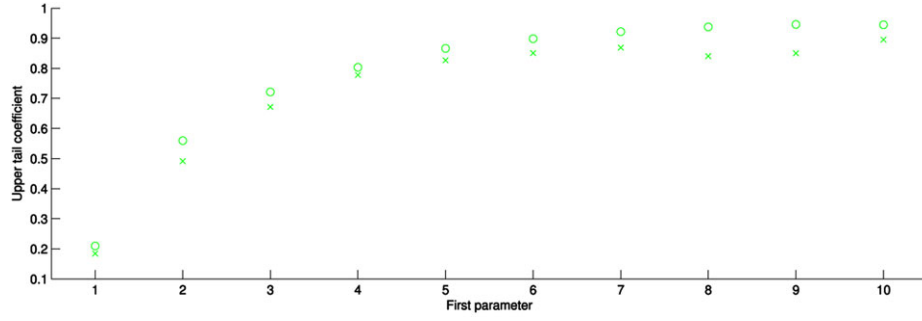


Fig. 13. Comparison of upper tail dependence coefficient for simulated values from Gumbel (green) copula and the minimum information D-vine. The circles represent the parametric copula and the crosses the minimum information vine in each case.

While the choice of vine structure imposes no restrictions on the underlying joint probability distribution it represents, the fact that we only use finite parameter families of copulas means that not every distribution is well represented, and that the choice of vine structure could be significant in closely matching a distribution while still using a simple family of copulas.

Any functions can be used to create the minimum information copulas used here, and in some applications it may be natural to use functions that are themselves computed in computer codes. Because of the frequent evaluation calls needed to determine the min inf distribution, it then makes sense to use emulators or Kriging models as a way to speed up the computations.

Finally, the methods used here generalize well-known methods such as *Normal to Anything*, used in simulation and decision analysis. This generalization provides us two main advantages: natural ways to generate asymmetric copulas, and simple ways to specify nonconstant conditional correlations (or other moments). Our methods provide a flexible methodology that may be adapted (through the choice of functions h_i and the choices about conditional expectations) to produce models taking account of naturally available information at an appropriate level of complexity for the modeling problem.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of EPSRC Grant EP/E018084/1 in supporting the research reported in this article. We thank the reviews and the associate editor for helpful comments.

APPENDIX A

Proof of Theorem 2

We have assumed that our multivariate density $f_{1\dots n}$ is a continuous function defined on $[0, 1]^n$. Since all marginal densities $f_{i_1\dots i_r}$ are obtained by integrating out variables from $f_{1\dots n}$, it is clear that:

$$|f_{i_1\dots i_r}(x_{i_1}, \dots, x_{i_r})| \leq \sup |f_{1\dots n}(x_1, \dots, x_n)|,$$

where the sup is taken over the variables x_i ($i \neq i_1, \dots, i_r$). Hence,

$$|f_{i_1\dots i_r}(x_i|x_{i_1}, \dots, x_{i_r})| = \left| \frac{f_{ii_1\dots i_r}(x_i x_{i_1}, \dots, x_{i_r})}{f_{i_1\dots i_r}(x_{i_1}, \dots, x_{i_r})} \right| \leq \|f\|/\alpha,$$

where $\alpha > 0$ is a lower bound on the values taken by f . This shows that there is a point-wise bound for all the functions in $\mathcal{M}(f)$.

In order to show equicontinuity, we first note that each function $f_{i_1\dots i_r}$ is uniformly continuous. Since there are only a finite number of such functions, we can always ensure that given $\epsilon > 0$ there is a $\delta > 0$ such that for any $i_1 \dots i_r$ if

$$|(x_{i_1}, \dots, x_{i_r}) - (y_{i_1}, \dots, y_{i_r})| < \delta,$$

then

$$|f_{i_1\dots i_r}(x_{i_1}, \dots, x_{i_r}) - f_{i_1\dots i_r}(y_{i_1}, \dots, y_{i_r})| < \epsilon \cdot \alpha.$$

Hence if $|x_i - y_i| < \delta$, then

$$|f_{i_1\dots i_r}(x_i|x_{i_1}, \dots, x_{i_r}) - f_{i_1\dots i_r}(y_i|x_{i_1}, \dots, x_{i_r})| \leq |f_{ii_1\dots i_r}(x_i, x_{i_1}, \dots, x_{i_r}) - f_{ii_1\dots i_r}(y_i, x_{i_1}, \dots, x_{i_r})|/\alpha \leq \epsilon,$$

and

$$\left| \frac{d}{d_{x_{i_k}}} f_{i_1\dots i_r}(x_{i_1}, \dots, x_{i_r}) \right| \leq \sup$$

so that $\mathcal{M}(f)$ must also be an equicontinuous family. A similar argument shows that $\mathcal{B}(f)$ is an equicontinuous family.

Proof of Theorem 3

For any element $c_{ij|i_1 \dots i_r}$ of $\mathcal{C}(f)$, we have

$$\begin{aligned} c_{ij|i_1 \dots i_r}(u_i, u_j | x_{i_1} \dots x_{i_r}) \\ = \frac{f_{ij|i_1 \dots i_r}(x_i, x_j | x_{i_1} \dots x_{i_r})}{f_{i|i_1 \dots i_r}(x_i | x_{i_1} \dots x_{i_r}) f_{j|i_1 \dots i_r}(x_j | x_{i_1} \dots x_{i_r})}. \end{aligned}$$

Hence if we take a sequence of elements in $\mathcal{C}(f)$, then there are corresponding sequences of elements of $\mathcal{M}(f)$ and $\mathcal{B}(f)$. Since $\mathcal{M}(f)$ is relatively compact there must be a convergent subsequence, and looking along that same subsequence there must be a subsequence of that for which the corresponding functions in $\mathcal{B}(f)$ converge. Now, along this subsequence the right-hand side of the above expression converges, so the elements of $\mathcal{C}(f)$ on this same sequence must converge (and to the same thing). In particular, there is a convergent subsequence. Hence, $\mathcal{C}(f)$ is relatively compact.

Proof of Lemma 1

We show that by taking f sufficiently close to g one can ensure that the reweighting functions for f are close to 1. This then implies that $C(f)$ is close to g .

Without loss of generality we can assume that f is normalized. The proof uses the fact that we can use the Borwein-Lewis-Nussbaum approach to find functions $d_{1f}(u)$ and $d_{2f}(v)$ such that $d_{1f} \cdot d_{2f} \cdot f$ has uniform marginals. Such functions d_{1g} and d_{2g} exist

also for g but are constant, $d_{1g}(u) = d_{2g}(v) = 1$, because g is already a copula. As discussed above, these reweighting functions are fixed points of a functional that is a contraction mapping when using the Hilbert metric \mathcal{D} on the appropriate space of pairs of functions (d_1, d_2) .

We denote by L_f the functional associated to f . Since this is a contraction mapping there exists a $\lambda_f \in (0, 1)$ such that

$$\mathcal{D}(L_f(a, b), L_f(c, d)) < \lambda_f \mathcal{D}((a, b), (c, d)).$$

If we set $a_0 = 1, b_0 = 1$, and $(a_{n+1}, b_{n+1}) = L_f(a_n, b_n)$, then we have convergence to the required pair of functions (d_{1f}, d_{2f}) that reweight f to become a copula.

Now, by choosing f close enough to g we can ensure two things. First, that the contraction rate associated to L_f is close to that of L_g , in particular less than some chosen $\lambda < 1$. Second, we can ensure that

$$\mathcal{D}(L_g(1, 1), L_f(1, 1)) = \mathcal{D}((1, 1), L_f(1, 1))$$

is as small as required. This implies that

$$\begin{aligned} \mathcal{D}((1, 1), (a, b)) &\leq \sum_{n=0}^{\infty} \mathcal{D}((a_n, b_n), (a_{n+1}, b_{n+1})) \\ &\leq \mathcal{D}((a_0, b_0), (a_1, b_1)) \sum_{n=0}^{\infty} \lambda^n \\ &= \frac{\mathcal{D}((1, 1), L_f(1, 1))}{1 - \lambda}. \end{aligned}$$

Hence, the reweighting functions for f are close to the identity, and so $C(f)$ is close to g .

APPENDIX B

8. Jouini MN, Clemen RT. Copula models for aggregating expert opinions. *Operations Research*, 1996; 44:444–457.

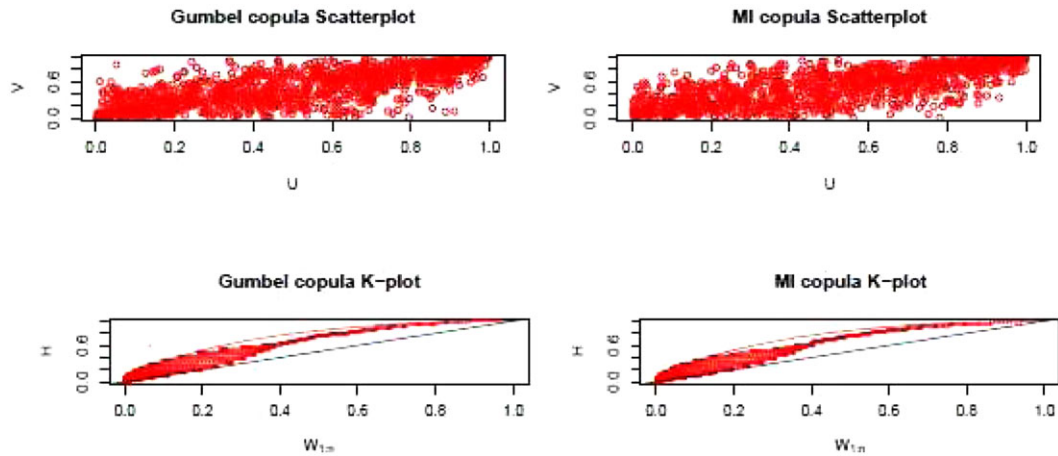


Fig. B.1. Scatter- and K-plots of the Gumbel copula and the fitted minimum information copula.

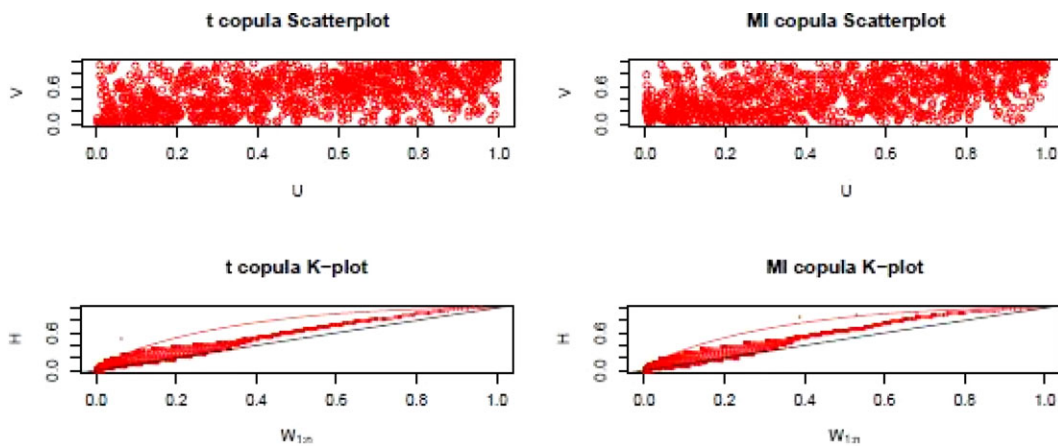


Fig. B.2. Scatter- and K-plots of the t -copula and the fitted minimum information copula.

REFERENCES

1. Nelsen RB. Copulas, characterization, correlation and counterexamples. *Mathematics Magazine*, 1995; 68:193–198.
2. Nelsen RB. *An Introduction to Copulas*. New York: Springer-Verlag, 1999.
3. Joe H. *Multivariate Models and Dependence Concepts*. London: Chapman & Hall, 1997.
4. Abbas AE. Multiattribute utility copulas. *Operations Research*, 2009; 57:1367–1383.
5. Abbas AE, Budescu DV, Gu YH. Assessing joint distributions with isoprobability contours. *Management Science*, 2010; 56:997–1011.
6. Biller B. Copula-based multivariate input models for stochastic simulation. *Operations Research*, 2009; 57:878–892.
7. Clemen RT, Reilly TH. Correlations and copulas for decision and risk analysis. *Management Science*, 1999; 45:208–224.
8. Kotz S, van Dorp JR. Generalized diagonal band copulas with two-sided generating densities. *Decision Analysis*, 2010; 7:196–214.
9. Ghosh S, Henderson SE. Behaviour of the NORTA method for correlated random vector generation as the dimension increases. *ACM Transactions on Modeling and Computer Simulation*, 2003; 13:276–294.
10. Aas K, Czado KC, Frigessi A, Bakken H. Pair-copula constructions of multiple dependence. *Insurance, Mathematics and Economics*, 2009; 44:182–198.
11. Deng L, Ma C, Yang W. Portfolio optimization via pair copula-GARCH-EVT-CVaR model. *Systems Engineering Procedia*, 2011; 2:171–181.
12. Bedford T, Cooke RM. Vines—A new graphical model for dependent random variables. *Annals of Statistics*, 2002; 30:1031–1068.

14. Kurowicka D, Cooke RM. Uncertainty Analysis with High Dimensional Dependence Modelling. Chichester: John Wiley, 2006.
15. Bedford T, Cooke RM. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 2001; 32:245–268.
16. Czado C. Pair-Copula Constructions of Multivariate Copulas. New York: Springer, 2010.
17. Kurowicka D, Joe H. Dependence Modeling. Vine Copula Handbook. Singapore: World Scientific Publishing, 2011.
18. Bauer A, Czado C, Klein T. Pair-copula constructions for non-Gaussian DAG models. *Canadian Journal of Statistics*, 2012; 40:86–109.
19. Dißmann J, Brechmann EC, Czado C, Kurowicka D. Selecting and estimating regular vine copulae and application to financial return. *Computational Statistics and Data Analysis*, 2013; 59:52–69.
20. Acar EF, Genest C, Neslehova J. Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, 2012; 110:74–90.
21. Haff I, Aas K, Frigessi A. On the simplified pair-copula construction—Simply useful or too simplistic? *Journal of Multivariate Analysis*, 2010; 101:1296–1310.
22. Stoeber J, Joe H, Czado C. Simplified pair copula constructions—Limitations and extensions. *Journal of Multivariate Analysis*, 2013; 119:101–118.
23. Lopez-Paz D, Hernandez-Lobato JM, Ghahramani Z. Gaussian process vine copulas for multivariate dependence. Pp. 10–18 in *Proceedings of the 30th International Conference on Machine Learning*, Vol. 28, 2013.
24. Shreve S. Did mathematical models cause the financial fiasco? *Analytics Magazine*, 2009, Spring, 2009; 6–7.
25. Jaynes ET. *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press, 2003.
26. Bedford T, Meeuwissen A. Minimally informative distributions with given rank correlation for use in uncertainty analysis. *Journal of Statistical Computation and Simulation*, 1997; 57:143–174.
27. Kurowicka D, Morales O, Roelen A. Eliciting conditional and unconditional rank correlations from conditional probabilities. *Reliability Engineering and System Safety*, 2008; 93:699–710.
28. Bedford T. Interactive expert assignment of minimally-informative copulae. *Applied Probability Trust*, 2006; 2:1–5.
29. Whittle P. *Probability via Expectation*. New York: Springer, 1970.
30. Lad F. *Operational Subjective Statistical Methods*. Chichester: John Wiley, 1996.
31. Goldstein M, Wooff D. *Bayes Linear Statistics: Theory and Methods*. Chichester: Wiley, 2007.
32. Min A, Czado C. Bayesian inference for multivariate copulas using pair-copula constructions. *Journal of Financial Econometrics*, 2010; 8:511–546.
33. Aven T. On the need for restricting the probabilistic analysis in risk assessments to variability. *Risk Analysis*, 2010; 30:354–360.
34. Pat  Cornell ME. Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering & System Safety*, 1996; 95–111.
35. Borgonovo E. Epistemic uncertainty in the ranking and categorization of probabilistic safety assessment model elements: Issues and findings. *Risk Analysis*, 2008; 28:983–1001.
36. Borwein J, Lewis A, Nussbaum R. Entropy minimization, DAD problems, and doubly stochastic kernel. *Journal of Functional Analysis*, 1994; 123:264–307.
37. Nussbaum RD. *Iterated Nonlinear Maps and Hilbert’s Projective Metric, II*. Providence: American Mathematical Society, 1989.
38. Bedford T, Wilson KJ. On the construction of minimum information bivariate copula families. *Annals of the Institute of Statistical Mathematics*, 2014; 66:703–723.
39. Lewandowski D. High Dimensional Dependence: Copulas, Sensitivity, Sampling. Ph.D. Thesis, Delft University, 2008.
40. Sinkhorn R, Knopp P. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 1967; 21:343–348.
41. Csiszar I. I-Divergence geometry of probability distributions and minimization problems. *Annals of Probability*, 1975; 3:146–158.
42. Lagarias JC, Reeds JA, Wright MH, Wright PE. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal of Optimization*, 1998; 9:112–147.
43. Salmon F. The formula that killed Wall Street. *Significance*, 2012; 9(1):16–20.
44. Ababei DA, Kurowicka D, Cooke RM. Uncertainty analysis with UNICORN. *Proceedings of the Third Brazilian Conference on Statistical Modelling in Insurance and Finance*, 2007.
45. Genest C, Boies JC. Detecting dependence with Kendall plots. *American Statistician*, 2003; 57:275–284.
46. Frahm G, Junker M, Schmidt M. Estimating the tail dependence coefficient: Properties and pitfalls. *Insurance: Mathematics and Economics*, 2005; 37:80–100.
47. Ganguly P, Reddy MJ. Probabilistic assessment of flood risks using bivariate copulas. *Theoretical and Applied Climatology*, 2013; 111:341–360.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher’s website:

Algorithm 1 To approximate the joint density between two variables of interest, X and Y , using a minimally informative copula.

Algorithm 2 To find the log-likelihood of a minimally informative copula between X and Y given data x_1, \dots, x_n and y_1, \dots, y_n .

Algorithm 3 Calculates the conditional distribution (F_c) or density (f_c) on a grid of points from joint density f .

Algorithm 4 Algorithm to simulate from a 4-dimensional D-vine in (X_1, X_2, X_3, X_4) given uniform marginals and minimum information copulas.