Copulas: Tales and facts

Thomas Mikosch

"But he does not wear any clothes" said the little child in Hans Christian Andersen's "The Emperor's New Clothes."

1 Some Preliminary Facts

When I started writing the paper (Mikosch, 2005) in 2003 a Google search of the word "copula" gave 10,000 responses. In September 2005 the same search gives 650,000 responses. There is an explosion of activity. What is going on?

Many of the web-sites[1] found in the Google search are related to mathematical finance, statistics, extreme value theory, and risk management. Everybody who opens any journal on stochastic processes, probability theory, statistics, econometrics, risk management, finance, insurance, etc., observes that there is a fast growing industry on copulas. The commercial statistics software Splus

---

[1] Of course, the word "copula" is also used in rather distinct non-mathematical contexts.

T. Mikosch (✉)
Laboratory of Actuarial Mathematics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark
e-mail: mikosch@math.ku.dk

provides the module FinMetrics that includes copula fitting which is written by R. Carmona, see also Carmona (2004). One can also get copula modules in other major software packages (R, Mathematica, MatLab, etc.). The International Actuarial Association (2004) in its hefty paper on Solvency II[2] recommends using copulas for modeling dependence in insurance portfolios. Moody's uses Gaussian copulas for modeling credit risk and provides software for it which is used in many financial institutions. Since Basle II[3] copulas are now standard tools in credit risk management.

The main purpose of this paper is to ask some naïve questions about the fast ascent of copula technology that has become so fashionable. My main concern is that this very simple concept might be something like the emperor's new clothes because it promises to solve all problems of stochastic dependence but it falls short in achieving the goal. Although I do appreciate that practitioners, in contrast to academic researchers, have to come up with solutions to their risk problems within deadlines and that "quick and dirty methods" cannot always be avoided. Yet one may of course ask how much safety the banking and insurance industry (and maybe the rest of the world) really gains by using the copula concept.

I have been watching my colleagues for some years and have been wondering why more and more of them became immersed in copulas without discussing the pros and cons of the concept. I suspect that some include the word "copula" in the title of their papers not because they contribute to the theory on copulas, but because they believe that one can publish easier. Some have adopted the language of copulas which has led them to publish papers with complicated technical assumptions, whereas the results are not new when considered in the usual language of distributions. I also observed that my students are more likely to be attracted to copulas than to stochastic processes. A possible reason is that one needs less than 10 minutes to understand the fundamentals of copulas, but many years of studies in order to get an idea of a genuine stochastic process.

I got the impression that some of those who use copulas or write about them think that all the world's problems related to stochastic dependence and multivariate distributions can be solved via copulas. It is my primary objective to caution this optimism. A second goal is to indicate that there are statistical problems in handling copulas, one of them being the curse of dimensionality. A third aim is to indicate that copulas do not really fit into the theory of stochastic processes and time series analysis. This is a pity because a new strong idea (if it is one) should inherit the main body of a well established theory in which

---

[2]Solvency II will be a treaty for insurers similar to the Basle I and II treaties for banks. The European Commission writes about Solvency II: *It contains a fundamental and wide-ranging review of the current regime in the light of current development in insurance, risk management, finance techniques, financial reporting, etc. One of the key objectives of Solvency II is to establish a solvency system that is better matched to the true risks of an insurance company.*

[3]See the website www.bis.org/publ/bcbsca.htm of the Bank for International Settlements for a guide to credit and market risk.

some of the finest minds of probability theory have been working for about 100 years.

I am aware that it is difficult to interfere with a development which is already driving at a very high speed, but an academic person without commercial interests may ask questions which do not make all people in the community happy. In what follows, I will ask questions about copulas and what one gains scientifically if one uses them.

2 What is a Copula?

We start with a real-valued random variable X with continuous distribution function F which we also assume to have an inverse F^{-1} on its support. It is easily checked that $X \stackrel{d}{=} F^{-1}(U)$ where U is uniformly distributed on $(0, 1)$ and $F(X) \stackrel{d}{=} U$ and $\stackrel{d}{=}$ refers to equality in distribution.

Now consider an $\mathbb{R}^d$-valued random vector $X = (X_1, \ldots, X_d)$ with respective marginal distribution functions F_i which are again assumed to be continuous and have inverse on the support of X_i.[4] Then the same arguments as in the one-dimensional case yield

$$P(X_1 \leq F_1^{-1}(x_1), \ldots, X_d \leq F_d^{-1}(x_d)) = P(F_1(X_1) \leq x_1, \ldots, F_d(X_d) \leq x_d)$$

$$= C(x_1, \ldots, x_d),$$

where on the right-hand side the d-dimensional distribution function C has support $[0, 1]^d$ and uniform marginal distribution functions. The distribution function C is the *copula* of the vector X. On the other hand, given the copula C of X we can reconstruct the distribution function of X by the distribution functions F_i:

$$P(X_1 \leq y_1, \ldots, X_d \leq y_d) = P(F_1(X_1) \leq F_1(y_1), \ldots, F_d(X_d) \leq F_d(y_d))$$

$$= C(F_1(y_1), \ldots, F_d(y_d)).$$

Finally, any distribution function with support on $[0, 1]^d$ and uniform marginals is called a *copula*.

Under the given assumptions on the marginal distribution functions F_i, there is a 1-to-1 correspondence between the copula C and the distribution of X. This means, in a sense, that the dependence structure of $X_1, \ldots, X_d$ can be *reconstructed* from the copula and the marginal distributions F_i. But this does not mean that the dependence structures of the vectors $X = (X_1, \ldots, X_d)$ with copula C and of the vector $Y = (F_1(X_1), \ldots, F_d(X_d))$ with distribution

---

[4]These assumptions can be weakened but this paper does not aim to discuss the most general setting. See Nelsen (1999) for an introduction to general copulas and their properties.

function C are the same. For example, if the correlations of the vector X are defined, they will usually be different from those of Y. To illustrate this fact more drastically, assume that $X_1, \ldots, X_d$ is a sample from a second order stationary process with zero autocorrelations. The corresponding sample $F_1(X_1), \ldots, F_d(X_d)$ can be highly correlated and will in general not be second order stationary. Similarly, the spectral measure (which is an extremal dependence measure, see Section 8.2) can make sense for X with unbounded support, but it does not for Y.

3 Why did Copulas Become Popular in Risk Management?

One of the driving forces for the popularity of copulas is their application in the context of financial risk management. Standard theory in mathematical finance prescribes Gaussian processes among others to log-prices and interest rates.[5] This is due to the dominating role of Brownian motion and martingales in models of mathematical finance. The distribution of a Gaussian process is determined by its finite-dimensional distributions, which are Gaussian and, in turn, are completely characterized by their mean values and covariances. In particular, a linear combination of the components of a mean zero Gaussian vector is Gaussian and its distribution is determined purely by its covariance matrix and the weights of the linear combination. This is convenient because in risk management one is often interested in linear combinations of prices or log-prices from different assets which constitute a portfolio. Quantities such as Value at Risk (VaR—a low quantile of the Profit and Loss distribution) can be calculated if one knows the covariance matrix and the number of shares one holds in the various assets.

Unfortunately, real-life data often turn out to be highly non-Gaussian, see for example Chapter 6 in Embrechts et al. (1997) or Mikosch (2003) for some evidence of Pareto like tails for returns. Returns (the increments of log-prices over successive periods of time with equal length, such as days) have heavy-tailed distributions, are dependent through time and may be non-stationary. There is no simple alternative to the Gaussian distribution in the non-Gaussian world. In particular, one needs multivariate models for portfolios with different marginal distributions (including different tail behavior) and a dependence structure which is determined not only by covariances. Many of the well known multivariate distributions are not flexible enough to allow for different tail behavior in different components. Therefore copulas seem to be the right tools in order to overcome the mentioned difficulties: they generate all multivariate distributions with flexible marginals in the way described in Section 2.

So far so good. But at least at this point any critical person might have some questions. For example the following one.

---

[5]For example, the celebrated Black–Scholes model assumes a Brownian motion; the popular Vasicek interest rate model requires an Ornstein–Uhlenbeck process.

4 Why does One Transform the Marginals to a Uniform Distribution?

The author has not found a *scientific* answer to this question. But my optimistic guess is that this transformation is related to the fact that one can visualize data from a two-dimensional copula on a computer screen which has some similarity with the unit square $[0, 1]^2$. Such two-dimensional visualizations have become rather standard as an exploratory tool. Another explanation might be that the transformation F^{-1} is well known as the quantile function of F and that the relationship between the uniform distribution, F and F^{-1} is part of any advanced probability course, hence very familiar.

However, what does such a visualization tell or *not tell us*? It would be useful if one could interpret the dependence structure of vectors with uniform marginal distributions. But there are infinitely many such dependencies and to the best of my knowledge there is no way of classifying or quantifying them. It is even hard to imagine what dependence of two uniformly distributed random variables means. Hence:

> *There is no particular mathematical or practical reason for transforming the marginals to the uniform distribution on* (0, 1).

With the same right, one could transform the marginal distributions of a vector to any other nice continuous distribution function. For example, again assuming X to have a continuous distribution function F, the transformation $-\log(1 - F(X))$ yields a standard exponentially distributed random variable:

$$P(-\log(1 - F(X)) \le x) = P(1 - e^{-x} \ge F(X)) = 1 - e^{-x}, \quad x > 0.$$

Or one can transform the marginal distributions to standard normal ones. Then one will in general deal with a multivariate distribution which is not Gaussian. Such examples are sometimes considered in a course on probability theory when one wants to explain the fact that the (Gaussian) marginals do not determine a multivariate (Gaussian) distribution. A multivariate non-Gaussian distribution with Gaussian marginal distributions is usually considered pathological and not of much practical use. But such a distribution corresponds to a copula. Would this copula not be "pathological" as well?

Later, in Section 7.2, we will discuss Gaussian and t-copulas which are generated from multivariate Gaussian and t-distributions, respectively. One wonders why, in this context, one does not transform the marginals to Gaussian or t-distributions. In contrast to the corresponding copulas the characteristics of these distributions have a straightforward interpretation.

5 Why does One Transform the Marginals to a Standard Distribution?

In various situations one can apply increasing transformations to the data without destroying certain aspects of the underlying dependence structure. This

can be advantageous if one gets simpler representations or more accessible formulas for certain probabilistic quantities. Multivariate extreme value theory is in part responsible for this approach. Standard multivariate extreme value theory deals with the vector of component-wise maxima of a sample of iid random vectors and its limit distribution under affine transformations.[6] If the components of these vectors come from different distributions, the question about the limit distribution of a vector of component-wise maxima can be rather unpleasant. For example, different marginal distributions require different normalizations or, if one chooses the normalization from a dominating component, the remaining components might vanish in the limit. Therefore it has become rather common to assume that one transforms the components to a standard distribution.[7] Those include the uniform distribution, but also the Pareto, the unit Fréchet distribution $\Phi_1(x) = e^{-1/x}$, $x > 0$, and the Gumbel distribution $\Lambda(x) = \exp\{-e^{-x}\}$, $x \in \mathbb{R}$. The advantage of choosing the Pareto or the unit Fréchet distributions is that they allow for an interpretation of the spectral measure, i.e., the distribution of the directions of multivariate extremes (see Section 8.2) given one believes that, after the transformation to these standard marginal distributions, the vector has a distribution in the maximum domain of attraction (MDA) of a suitable multivariate extreme value distribution (EVD).[8]

Increasing transformations of the marginals make sense in the context of multivariate extreme value theory. For example, consider the iid $\mathbb{R}^2$-valued vectors (X_i, Y_i) and transform the first and second components by increasing functions f_1 and f_2, respectively. Then the components of the transformed vectors $(f_1(X_i), f_2(Y_i))$ have the same ordering as before, and with the back transformations f_1^{-1} and f_2^{-1} one can reconstruct the distribution of the original vectors (X_i, Y_i). Of course, apart from keeping the order of the components unchanged by these monotone transformations, one destroys many other aspects of the dependence structure of the points (X_i, Y_i).

For many other functions of random vectors increasing transformations of the components do not really help one to evaluate the distribution of this function. For example, when one is interested in VaR, see Section 3, i.e., in a

---

[6]This is what the layman thinks about multivariate extreme value theory and he is right when he reads most of the literature, including textbooks. But component-wise maxima of a multivariate sample give a rather limited picture of the world of multivariate extremes. In my view, it is more important that the theory provides maximum domain of attraction (MDA) conditions which justify the assignment of probabilities to multivariate rare events which, possibly, have not happened before.

[7]This approach is advocated in Resnick's monograph (Resnick, 1987), where he lays the probabilistic foundations of multivariate extreme value theory. It is done by an easy probabilistic argument, but when dealing with data this is hard statistical work because one does not know the marginal distributions; see e.g. de Haan and Ronde (1998) and the literature cited therein. The problem is the same when one wants to verify the goodness-of-fit of a copula.

[8]As most conditions in extreme value statistics these are statistical believes which can hardly be tested on data.

certain quantile of a linear combination of the components of a random vector, increasing transformations of those components are not useful: the copula of the vector does not help one to evaluate VaR.

However, if one believes in a particular copula C and in particular marginal distributions F_i, then one can possibly simulate iid copies from the underlying distribution and, in turn, approximate the distribution of the function by crude Monte Carlo. This method has attracted some attention in risk management because it gives one a convenient Monte Carlo method for determining the VaR of a portfolio for a flexible number of shares (portfolio weights).

Of course, this Monte Carlo method also works if one simulates directly from a multivariate distribution without knowing its copula. Copulas do not solve the problem of simulating from an arbitrary multivariate distribution. Since one does not understand all dependence structures of a vector with values in $[0, 1]^d$ and uniform marginals, one cannot simulate from any copula C. Most copulas from which one can simulate are derived from well known multivariate distributions such as the Gaussian or elliptical distributions (see Section 7.2) and then one can simulate directly from these distributions.

The crude Monte Carlo method for simulating the quantile VaR is a rather expensive one. One has to simulate from the whole multivariate distribution in order to calculate the quantile of a one-dimensional object—a linear combination of the components of the vector. One may wonder whether there do not exist theoretical means for approximating the VaR for a sufficiently large class of high-dimensional distributions without knowledge of the distribution in its whole domain. Or formulated as a question: given a high-dimensional random vector, which part of its support is responsible for the far out upper/lower tail of a linear combination of its components? For elliptical distributions (see Section 7.2) one has explicit formulas for VaR, depending only on the parameters of the distributions. This fact contributed to the popularity of these distributions.

In a way, the distribution of a linear combination of a high-dimensional vector of prices and of a composite stock index (such as the Dow Jones, S&P 500, DAX, etc.) are similar objects; composite stock indices are weighted averages of a large number of speculative prices. Over the last 20 years econometrics has quite successfully developed time series models for returns of speculative prices and composite stock indices such as the ARCH family and stochastic volatility models. This development was crowned by two Nobel Prizes for Economics in 2003. Weighted averages of prices have their own dynamics and lead to interesting space-time dependence structures. If one believes in the feasibility of the copula approach one might wonder whether it would yield a better approximation to the marginal distribution of the returns of a composite stock index than what is obtained by fitting a time series model. To the best of my knowledge such a comparison has not been provided.

It is not only the ordering of a sample that is preserved under increasing transformations. Various quantities of interest are invariant under increasing transformations. Among those are some quantities which depend only on two-dimensional distributions and which are considered as some kind of

dependence measure: the *tail dependence parameter* (see Section 8.1), the concordance measures *Kendall's* τ and *Spearman's* ρ (see Nelsen (1999), Section 5.1). One says that two points (a, b) and (c, d) in $\mathbb{R}^2$ are *concordant* if $(a - c)(b - d) > 0$ and *discordant* otherwise. If one transforms the first or second coordinates by increasing functions, the resulting points remain concordant or discordant. Since τ and ρ measure the degree of concordance of a two-dimensional random vector they remain unchanged under increasing transformations of its components and only depend on the copula.

The mentioned concordance measures do not characterize the dependence of a two-dimensional vector with uniform marginals; they describe a very particular geometric property of an iid or strictly stationary ergodic sequence of two-dimensional random vectors. The *rank correlations* τ and ρ are often mentioned in the literature on risk management as alternative measures to covariances and correlations. They are not relevant for measuring dependence *and* risk because they do not depend on the magnitude of the data and therefore they neglect large and small values.

In the history of probability theory a whole universe of dependence measures of random vectors, stochastic processes and their distributions has been developed which describe very different aspects of stochastic dependence. Among them is the classical theory of probability metrics and distances as presented in the work of Zolotarev and summarized in his monograph (Zolotarev, 1997). As a matter of fact, there has always been some interest in a probabilistic theory of distributions with fixed marginals; for recent contributions see the work by Rüschendorf and his co-authors.

6 How does One Fit the Marginal Distributions and the Copula?

In classical statistics one fits a multivariate distribution by extracting information out of the data about the distribution as an entity, e.g., by using maximum likelihood for a multivariate parametric family of distributions. The copula technology is different since it suggests the possibility of a two stage statistical procedure: estimate the marginal distributions and the copula function separately from each other. By the choice of the marginal distributions one determines the copula, hence the chosen dependence structure, and therefore different statistical tools for fitting the marginals may generate distinct dependencies.

In most parts of the existing literature the question about the fit of the marginals is not discussed. An exception is Section 5.5 in McNeil et al. (2005) where three possible approaches are mentioned:

(1)   Fit parametric distributions to the marginals.
(2)   Replace the marginal distributions F_i by their empirical distributions $F_i^{(n)}$.
(3)   Replace the marginal distributions F_i by their empirical counterparts $F_i^{(n)}$ and add some tails, e.g., generalized Pareto distribution (GPD) tails from above or below a certain threshold.

From one-dimensional statistical theory it is known that any of these approaches can go terribly wrong. For example, if one wants to estimate high or low quantiles the empirical distribution can be a poor guide, even when the sample size is large. If one estimates the marginal distributions by its empirical counterparts the estimation of the copula is entirely based on the relative ranks of the components: the data can be multiplied or divided by 10^{50} and the statistical analysis will be the same. *Can the estimation of the dependence structure only be a matter of relative ranks?*

From a one-dimensional extreme value point of view the approximation of the tails of the marginal distributions by GPDs might be appropriate *if one is interested in an extreme value problem*, but the estimation of the copula depends on all data, not just on the extremes, even if the copula comes from an EVD.

Any of these approaches will only work under particular *mathematical conditions*. None of these methods can be trusted unless a rigorous mathematical proof is given. There exists very little statistical theory about fitting multivariate data by using copulas and about their goodness-of-fit; most papers focus on some particular two-dimensional copulas (Archimedean, Gaussian, *t*-copula, etc.); see e.g. Genest and Rivest (1993) or Song et al. (2003). For statistical applications it is not enough to prove that maximum likelihood estimation works for some parametric families of copulas and marginal distributions. A major problem is the goodness-of-fit of a copula which critically depends on the fit of the marginal distributions. In financial applications copulas are often used for high-dimensional data (e.g. Moody's credit risk models), and it is not clear what is the trade-off between the sample size and the dimension. The number of parameters might be larger than the sample size. Of course, the same problem occurs if one fits a Gaussian distribution in a naïve way to two years of return data, a sample of 500 points say, to a portfolio which consists of 200 assets, i.e., the return data are 200-dimensional. It is hard to believe that any copula technique will sweep away these statistical problems.

The separation of marginal distributions and dependence function when using distributions generated from a copula implies a much higher statistical uncertainty than usually encountered when fitting a distribution. Instead of fitting one multivariate distribution as an entity to the data, one has to fit all marginal distributions as well as the copula by using one sample. This increases the usual estimation efforts by far. It is unclear how sensitive the estimation techniques for the copula are to the estimation of the marginal distributions. It is also unclear how sensitive the estimation of the distribution of the multivariate data is to the estimation of the marginal distributions *and* the copula. Here is a natural question:

> *Given the fitted marginal distributions and copulas are "close" in some sense to their counterparts in the distribution of the data, how "close" is the estimated distribution (generated by the fitted marginal distributions and estimated copula) to the distribution of the data?*

7 How does One Choose a Copula?

The copula C of a random vector X is a transformation of the distribution of X. As for the distribution of X we have infinitely many choices for C. If one chooses a copula it should be related to the problem at hand. For example, if we are interested in multivariate extremes the copula should be related to multivariate extreme value theory; see Section 8 for some discussion in this context.

   In the literature various families of copula families with a name have been introduced. Their choice is not always based on reasoning but on mathematical convenience. In what follows, we comment on some of the popular copula families.

7.1 The Archimedean Copula

This copula has attracted a lot of attention, see e.g. Nelsen (1999), Chapter 4. In its simplest form it is defined as follows: let $\psi : [0, 1] \to [0, \infty]$ be a continuous decreasing function $\psi(0) = \infty$ and $\psi(1) = 0$. Moreover, for $d > 2$ the function ψ^{-1} on $[0, \infty]$ also has to be completely comonotonic.[9] Then the following function on $[0, 1]^d$ is a copula, see Nelsen (1999), Theorem 4.6.2 on p. 122:

$$C(x_1, \ldots, x_d) = \psi^{-1} \left(\sum_{i=1}^{d} \psi(x_i) \right).$$

Calculation yields

$$P(X_1 \leq x_1, \ldots, X_d \leq x_d) = C(F_1(x_1), \ldots, F_d(x_d)) = \psi^{-1} \left(\sum_{i=1}^{d} \psi(F_i(x_i)) \right)$$

$$= \psi^{-1} \left(- \log \left[P(Y_1 \leq x_1) \cdots P(Y_d \leq x_d) \right] \right),$$

where $Y = (Y_1, \ldots, Y_d)$ is a vector with independent components with distribution functions

$$P(Y_i \leq x) = e^{-\psi(F_i(x))}, \quad i = 1, \ldots, d.$$

This means that the distribution function of X is a monotone transformation of the distribution function of a vector Y with independent components.

---

[9]ψ^{-1} is completely monotonic on $[0, \infty]$ if $(-1)^k d^k \psi^{-1}(t)/dt^k \geq 0$ for all $t \in (0, \infty)$ and $k = 0, 1, 2, \ldots$; see Nelsen (1999), p. 122.

The Archimedean copula is a textbook toy example where one can explain various theoretical properties of a copula and calculate quantities such as the rank correlations. However, in view of the above calculation one may ask:

> *In which practical situation would one choose an Archimedean copula for modeling genuine* dependence *in the vector X?*

7.2 The Gaussian, *t*- and Other Copulas of Elliptical Distributions

The *Gaussian copula* is perhaps the most popular distribution in applications. It is simply derived from a multivariate Gaussian distribution function Φ_Σ with mean zero and correlation matrix Σ by transforming the marginals by the inverse of the standard normal distribution function Φ:

$$C(x_1, \ldots, x_d) = \Phi_\Sigma(\Phi^{-1}(x_1), \ldots, \Phi^{-1}(x_d)) \, .$$

The *t-copula* is another popular model. It is derived in the same way as the Gaussian copula. Given a multivariate centered *t*-distribution function $t_{\Sigma,\nu}$ with correlation matrix Σ, ν degrees of freedom and with marginal distribution function t_ν, this copula is given by

$$C(x_1, \ldots, x_d) = t_{\Sigma,\nu}(t_\nu^{-1}(x_1), \ldots, t_\nu^{-1}(x_d)) \, .$$

The Gaussian and *t*-copulas are copulas of *elliptical distributions*; they are not elliptical distributions themselves. A vector X with an elliptical distribution in $\mathbb{R}^d$ satisfies the identity in law $X \stackrel{d}{=} R A S$, where S is uniformly distributed on the unit sphere $\{x : |x| = 1\}$ with respect to the Euclidean norm $|\cdot|$, $R \geq 0$ is a radial random variable, independent of S, $AA' = \Sigma$ and Σ is a correlation matrix.

One may wonder how flexible and reasonable elliptical distributions and copulas are for the purposes of risk modeling. The multiplicative shock R is present in all components of X with the same magnitude. This would also imply that any observation X_i from such a distribution would be equally important for statistical analyses, independently of the order of magnitude of $|X_i|$. Such an approach is in contrast to the usual reasoning in extreme value statistics, where one first separates the data in the center of the distribution from those far away from zero and then conducts distinct statistical analyses in the different parts of the sample.

The vector $A S$ determines the likelihood of the directions of X. Since S is uniformly distributed on the unit sphere of $\mathbb{R}^d$, the distribution of $A S$ is also smoothly distributed in space; in particular extremes of an iid sample $X_1 = X, \ldots, X_n$ may occur in all directions whereas pronounced directions for extremes are excluded in these models.

The dependence in elliptical distributions is essentially determined by covariances. Covariances are rather poor tools for describing dependence for non-Gaussian distributions, in particular for their extremal dependence; see Embrechts et al. (2002) for a critique of using covariances in risk modeling and Glasserman (2004) for advocating *t*-distributions for risk management.

7.3 Extreme Value Copulas

An extreme value copula is derived from a multivariate EVD by transforming its marginals to the unit cube $[0, 1]^n$. Multivariate EVDs occur as weak limits of affinely transformed vectors of component-wise maxima of iid random variables; see Resnick (1987), Chapter 5. The marginal distributions are necessarily one-dimensional EVDs: they are either of the type Fréchet, Gumbel or Weibull.

A popular extreme value copula is the *Gumbel copula*

$$C(x_1, \ldots, x_d) = \exp \left\{ - \left(\sum_{i=1}^{d} (- \log x_i)^{1/\alpha} \right)^{\alpha} \right\} ,$$

for some $\alpha \in (0, 1]$. Writing $\Lambda(x) = \exp\{-e^{-x}\}$, $x \in \mathbb{R}$, for the Gumbel distribution function, the corresponding EVD is obtained by the transformation

$$C(\Lambda(x_1), \ldots, \Lambda(x_d)) = \exp \left\{ - \left(\sum_{i=1}^{d} e^{-x_i/\alpha} \right)^{\alpha} \right\} .$$

Other named (parametric) multivariate EVDs are given for example in Galambos (1987), Chapter 5, or in Kotz and Nadarajah (2000), Chapter 3. However, there exist infinitely many distinct multivariate EVDs with Gumbel, Fréchet or Weibull marginal distributions and therefore it is not quite clear why the Gumbel copula in particular has gained such popularity; it may perhaps be because it is also an Archimedean copula. As explained above, such copulas describe a rather limited kind of dependence structure of a random vector.

Extreme value copulas are not EVDs themselves, i.e., one loses the interpretation of an EVD as a max-stable distribution. EVDs should be fitted to data only if the data are generated by an "extremal mechanism", i.e., if they can be interpreted as component-wise maxima of iid random vectors. In the one-dimensional case, the *method of annual maxima* (see e.g., Embrechts et al. (1997), p. 317) is known to create such maxima. The method straightforwardly extends to higher dimensions. However, in the context of risk management it is less likely that, for example, daily return data of speculative prices are generated by an extremal mechanism, i.e., it is difficult to interpret them as maxima. For example, the use of extreme value copulas for the calculation of functionals such as VaR is out of context.

8 Upper Tail Dependence and Multivariate Extremes

If one browses on the Internet and searches for papers on copulas, one often finds arguments for discriminating between different distributions based on the value of the upper tail dependence parameter. For example, "the *t*-copula allows for *upper tail dependence* whereas the Gaussian copula does not and *therefore the t-copula is better for modeling dependence of multivariate*

extremes" or "the upper tail parameter of the Gumbel copula is smaller than that of the *t*-copula and therefore the *t*-copula models *stronger* extremal dependence". In what follows, we will shed some light on these statements.

8.1 The Tail Dependence Parameter

We start by defining the upper tail dependence parameter:[10] given a two-dimensional vector (X, Y) with $X \stackrel{d}{=} Y$ and right endpoint x_F of the underlying distribution, assume that the following limit exists

$$\lambda = \lim_{u \uparrow x_F} P(Y > u \mid X > u).$$

The parameter λ is called the *upper tail dependence parameter*. If $\lambda > 0$ then X and Y are *upper tail dependent* and asymptotically independent otherwise. If X and Y do not have identical distributions, one can also define λ via the copula C of (X, Y):

$$\lambda = \lim_{u \uparrow 1} \frac{1 - 2u + C(u, u)}{1 - u} = 1 + \lim_{u \uparrow 1} \frac{C(u, u) - u}{1 - u}, \qquad (8.1)$$

although it is not clear what one gains if one does this. In what follows, we will assume that $X \stackrel{d}{=} Y$.

A sufficient condition for the existence of λ is *regular variation* of $(f(X), f(Y))$ for some increasing transformation f; see Section 8.2 for a description and Resnick (1986, 1987, 2004) for a rigorous definition of multivariate regular variation. It can be relaxed to the weaker notion of *hidden regular variation*, see Heffernan and Resnick (2005), Maulik and Resnick (2004).

The quantity λ lies between zero (e.g. when X and Y are independent) and 1 (e.g., when $X = Y$). It is is not very informative as regards the joint extremal behavior of the vector (X, Y). In particular, it is restricted to the two-dimensional case. Of course, in certain circumstances one can calculate the dependence parameters λ_{ij} in a pairwise fashion between the components X_i and X_j of the vector $X = (X_1, \ldots, X_d)$, but these numbers do not determine the (extremal) dependence structure of the vector; for any set of upper tail dependence parameters one can find infinitely many vectors with the same parameters λ_{ij}.[11] Since the tail dependence parameter can be expressed only by the copula without knowing the marginals, see Eq. 8.1, the value λ is not a genuine extreme value concept which would have to take into account both the magnitude and the direction of extremes.

---

[10]The lower tail dependence parameter $\lim_{u \downarrow 0} P(Y \leq u \mid X \leq u)$ is defined in an analogous way.

[11]Assuming a multivariate regularly varying vector X, this claim follows from the representation Eq. 8.4 of λ.

8.2 The Maximum Domain of Attraction (MDA) of the Multivariate Fréchet Distribution

The upper tail dependence parameter describes only one aspect of the problem of extremal dependence, and it is restricted to the two-dimensional case. By its definition, the parameter λ measures the relative deviation of the probabilities $P(X > u, Y > u) = P((X, Y) \in u (1, \infty]^2)$ from the probability $P(X > u) = P((X, Y) \in u (1, \infty] \times [0, \infty])$ for large u. For an $\mathbb{R}^d$-valued vector X it is much more informative to consider the relative deviation of the probability $P(X \in u A)$ from the probability $P(X \in u B)$ for any two sets A and B bounded away from zero and for large u. This includes the two-dimensional case with $A = (1, \infty]^2$, $B = (1, \infty] \times [0, \infty]$.

If B is fixed—it is common to choose the set

$$B = \{x : |x| > 1\} \quad \text{for a given norm } | \cdot |; \tag{8.2}$$

a reasonable assumption (given the components of X have unbounded support) for conducting extreme value analysis is to require that the limits

$$\lambda_B(A) = \lim_{u \to \infty} \frac{P(X \in uA)}{P(X \in uB)} \tag{8.3}$$

exist for smooth[12] sets A bounded away from zero and are non-zero for some $A \neq B$. The quantities $\lambda_B(A)$ constitute a measure λ_B on the Borel sets of $(\mathbb{R} \cup \{\infty, -\infty\})^d \setminus \{0\}$ and one necessarily has $\lambda_B(tA) = t^{-\alpha} \lambda_B(A)$, $t > 0$, for some $\alpha > 0$, see Hult and Lindskog (2005). The condition Eq. 8.3 is referred to as *(multivariate) regular variation*. It is slightly more general than the condition of the same name used as the MDA condition for multivariate EVDs with Fréchet marginals $\Phi_\alpha(x) = e^{-x^{-\alpha}}$, $x > 0$, where one assumes that X has non-negative components; see Resnick (1986, 1987, 2004).

For the two-dimensional vector (X, Y) the upper tail dependence parameter λ is obtained by specifying $A = (1, \infty]^2$ and $B = (1, \infty] \times [0, \infty]$, hence

$$\lambda = \lambda_{(1,\infty] \times [0,\infty]}((1, \infty]^2). \tag{8.4}$$

In contrast to the measure λ_B, this number gives a rather limited view at two-dimensional extremes which are required to assume values in $(u, \infty]^2$ for large u. For example, the existence of the limit $\lambda > 0$ does not imply that multivariate extremes have a limit distribution under affine transformations, i.e., in general the distribution of (X, Y) is not in the MDA of a two-dimensional EVD because the considered events $\{(X, Y) \in (u, \infty]^2\}$ for large u contain rather restricted information about the location of extremes in $\mathbb{R}^2$.[13]

---

[12]This means that $\lambda_B(\partial A) = 0$.

[13]This means that the sets $(u, \infty]^2$ do not generate the vague convergence relation Eq. 8.3 on the Borel σ-field of $[0, \infty]^2 \setminus \{0\}$.

If we specify the set B as in Eq. 8.2, a measure theoretic argument shows that multivariate regular variation is equivalent to the existence of the limits

$$t^{-\alpha} G(S) = \lim_{u \to \infty} \frac{P(|X| > tu, X/|X| \in S)}{P(|X| > u)}, \quad t > 0,$$

for smooth[14] sets $S \subset \{s : |s| = 1\}$ and a probability distribution G on $\{s : |s| = 1\}$. This distribution is called the spectral measure of X. It has interpretation as the limit distribution of the direction $X/|X|$ of X, conditional on $|X| > u$ as $u \to \infty$.

Relation Eq. 8.3 offers a naïve means for approximating probabilities of rare events. For example, for positive t and an iid or weakly dependent stationary sample $X_1 = X, \ldots, X_n$,

$$P(X \in tuA) \sim t^{-\alpha} \lambda_B(A) \, P(|X| > u)$$

$$\approx t^{-\widehat{\alpha}} \widehat{\lambda}_B(A) \, n^{-1} \sum_{i=1}^{n} I_{(u,\infty)}(|X_i|) . \tag{8.5}$$

The right-hand expressions in Eq. 8.5 are estimators of the corresponding deterministic quantities. This simple-minded approach is similar to the GPD approximation by peaks over thresholds in the one-dimensional case. In addition to the one-dimensional difficulties (trade-off between bias and variance of the estimators depending on the threshold u) the complexity of estimation in higher dimensions increases enormously since one also has to estimate the measure λ_B (or the corresponding spectral measure).

8.3 Are Copulas Suitable for Modeling Multivariate Extremes?

Copulas generate any multivariate distribution. If one wants to make an honest analysis of multivariate extremes the distributions used should be related to extreme value theory in some way. In Section 7.3 we have briefly discussed extreme value copulas which correspond to multivariate EVDs. They are not appropriate, unless one believes that the underlying data are generated by some extremal mechanism. For financial or insurance data this assumption is questionable.

An alternative approach may be based on distributions in the MDA of an EVD. Then the data are not necessarily created by an extremal mechanism; see Section 8.2. Such an approach requires estimation of the measure λ_B or of the spectral measure G. These are not easy tasks, but they reflect the difficulty of multivariate analyses in general. In particular, large sample sizes are required. Statistical analyses have been performed and statistical theory has been provided in the 2- and 3-dimensional cases; see de Haan and Resnick (1993), de Haan and Ronde (1998), Einmahl et al. (2001). Applications of the

---

[14]This means that $G(\partial S) = 0$.

statistical theory for copulas (see e.g. Genest and Rivest (1993) or Song et al. (2003)) are also restricted to low-dimensional settings. The statistical problems mentioned will not be swept away by using copulas.

In some parts of the literature the notion of *copula domain of attraction* has been coined, see e.g. p. 315 in McNeil et al. (2005). Assume $Y_1, \ldots, Y_n$ is an iid sample whose common distribution is a copula C. Write M_n for the vector of component-wise maxima of those vectors and C_n for the copula of M_n. Then C is in the *copula domain of attraction* of some extreme value copula $\widetilde{C}$ if $C_n(x_1, \ldots, x_d) = C^n(x_1^{1/n}, \ldots, x_d^{1/n})$ converges weakly to $\widetilde{C}$ ($C_n \overset{w}{\to} \widetilde{C}$). This notion is redundant since it is equivalent to the concept of the MDA of a multivariate EVD. To see this, transform the marginals of C e.g., to unit Fréchet ones and denote the resulting multivariate distribution by F. Then $C_n \overset{w}{\to} \widetilde{C}$ is equivalent to the fact that F is in the MDA of a multivariate EVD with unit Fréchet marginals; see Galambos (1987), Theorem 5.2.3. There is no reason to introduce copula domains of attraction; they only create theoretical confusion and one does not gain new information about multivariate extremes.

9 How do Copulas Fit into the Models of Time Series Analysis and the Theory of Stochastic Processes?

Consider one of the standard time series models such as FARIMA, linear processes, GARCH, etc.; see Brockwell and Davis (1991, 1996) and Mikosch (2003) for definitions and properties of these models. Then it is in general impossible to say anything about the distribution of the lagged vector $(X_1, \ldots, X_n)$. Well known exceptions are linear processes with Gaussian or infinite variance stable innovations.

In general we do not even know the marginal distributions of a GARCH or an ARMA process and it can even be hard work to find asymptotic expressions for the tails of these distributions. In the case of regularly varying tails for linear processes (including ARMA) under the weakest possible conditions, see Mikosch and Samorodnitsky (2000), and Basrak et al. (2002) for power law tails of GARCH processes.

The same statement can be made about any standard stochastic process such as Markov process, martingale, diffusion, Lévy process, random measure, any interesting process which has been considered in the history of stochastic processes. Copulas completely fail[15] in describing complex space-time dependence structures. Their focus is on spatial dependence and the related statistics (mostly parametric maximum likelihood) are aimed at iid data. It is contradictory that in risk management, where one observes a lot of dependence through time, copulas are applied most frequently.

---

[15]Several authors have introduced *dynamic* copula models. I wonder how they incorporate standard time series or stochastic process models in their calculus.

10 Do Copulas Avoid the Curse of Dimensionality?

Since a copula is only a reformulation of the original distributional problem one cannot expect that a copula does better than any other high-dimensional model. Copulas will usually even increase the number of parameters to be fitted. For example, for the *t*-distribution one needs to estimate the underlying correlation matrix Σ and the degrees of freedom $\nu > 0$; see Section 7.2. For a model generated by the *t*-copula one needs in addition to estimate the parameters of the underlying marginal distributions.

If one has deeper insight into the underlying dependence structure of the data one can try to reduce the dimension d of a parametric model (such as the Gaussian copula) by assuming functional dependencies between the parameters (e.g. the covariances), or one might want to use dimension reduction techniques. The problem of reducing dimensions for distributions related to multivariate extreme value theory is an important unsolved problem: why should one deal with the extremes in all components of a multivariate sample if some of the components dominate the others?

11 The Copula Fashion—The Emperor's New Clothes?

Here are some final remarks which summarize my opinion about copulas.

- There is no particular advantage of using copulas when dealing with multivariate distributions. Instead one can and should use any multivariate distribution which is suited to the problem at hand and which can be treated by statistical techniques.
- The marginal distributions and the copula of a multivariate distribution are inextricably linked. The main selling point of the copula technology—separation of the copula (dependence function) from the marginal distributions—leads to a biased view of stochastic dependence, in particular when one fits a model to the data.
- Various copula models (Archimedean, *t*-, Gaussian, elliptical, extreme value) are mostly chosen because they are mathematically convenient; the rationale for their applications is murky.
- Copulas are considered as an alternative to Gaussian models in a non-Gaussian world. Since copulas generate any distribution the class is too big to be understood and to be useful.
- There is little statistical theory for copulas. Sensitivity studies of estimation procedures and goodness-of-fit tests for copulas are unknown. It is unclear whether a good fit of the copula of the data yields a good fit to the distribution of the data.
- Copulas do not contribute to a better understanding of multivariate extremes.
- Copulas do not fit into the existing framework of stochastic processes and time series analysis; they are essentially static models and are not useful for modeling dependence through time.

References

Basrak, B., Davis, R.A., Mikosch, T.: Regular variation of GARCH processes. Stoch. Process. Appl. **99**, 95–116 (2002)

Brockwell, P.J., Davis, R.A.: Time Series: Theory and Methods. 2nd edn. Springer, Berlin Heidelberg New York (1991)

Brockwell, P.J., Davis, R.A.: Introduction to Time Series and Forecasting. Springer, Berlin Heidelberg New York (1996)

Carmona, R.A.: Statistical Analysis of Financial Data in S-PLUS. Springer, Berlin Heidelberg New York (2004)

Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling Extremal Events for Insurance and Finance. Springer, Berlin Heidelberg New York (1997)

Embrechts, P., McNeil, A., Straumann, D.: Correlation and dependence in risk management: Properties and pitfalls. In: Dempster, M.A.H. (ed.) Risk Management: Value at Risk and Beyond, pp. 176–223. Cambridge University Press, UK (2002)

Glasserman, P.: Monte-Carlo Methods in Financial Engineering. Springer, Berlin Heidelberg New York (2004)

Galambos, J.: Asymptotic Theory of Extreme Order Statistics. 2nd edn. Krieger, Malabar, Florida (1987)

Genest, C., Rivest, L.-P.: Statistical inference procedures for bivariate Archimedean copulas. J. Am. Stat. Assoc. **88**, 1034–1043 (1993)

de Haan, L., Resnick, S.I.: Estimating the limit distribution of multivariate extremes. Commun. Stat., Stoch. Models **9**, 275–309 (1993)

de Haan, L., de Ronde, J.: Sea and wind: multivariate extremes at work. Extremes **1**, 7–45 (1998)

Heffernan, J., Resnick, S.: Hidden regular variation and the rank transform. Adv. Appl. Probab. **37**, 393–414 (2005)

Hult, H., Lindskog, F.: Extremal behavior of regularly varying stochastic processes. Stoch. Process. Appl. **115**, 249–274 (2005)

The International Actuarial Association: A global framework for insurer solvency assessment. Research Report. www.actuaries.org (2004)

Kotz, S., Nadarajah, S.: Extreme Value Distributions: Theory and Applications. Imperial College Press, London (2000)

Maulik, K., Resnick, S.: Characterizations and examples of hidden regular variation. Technical Report. www.orie.cornell.edu/~sid/ (2004)

McNeil, A., Frey, R., Embrechts, P.: Quantitative Risk Management: Concepts, Techniques, and Tools. Princeton University Press, New Jersey (2005)

Mikosch, T.: Modeling dependence and tails of financial time series. In: Finkenstädt, B., Rootzén, H. (eds.) Extreme Values in Finance, Telecommunications, and the Environment, pp. 187–286. Chapman & Hall, Boca Raton (2003)

Mikosch, T.: How to model multivariate extremes if one must? Stat. Neerl. **59**, 324–338 (2005)

Mikosch, T., Samorodnitsky, G.: The supremum of a negative drift random walk with dependent heavy-tailed steps. Ann. Appl. Probab. **10**, 1025–1064 (2000)

Nelsen, R.B.: An Introduction to Copulas. Lecture Notes in Statistic vol. 139, Springer, Berlin Heidelberg New York (1999)

Einmahl, J.H.J., de Haan, L., Piterbarg, V.I.: Nonparametric estimation of the spectral measure of an extreme value distribution. Ann. Stat. **29**, 1401–1423 (2001)

Resnick, S.I.: Point processes, regular variation and weak convergence. Adv. Appl. Probab. **18**, 66–138 (1986)

Resnick, S.I.: Extreme Values, Regular Variation, and Point Processes. Springer, Berlin Heidelberg New York (1987)

Resnick, S.I.: On the foundations of multivariate heavy-tail analysis. Stochastic methods and their applications. J. Appl. Probab. **41**A, 191–212 (2004)

Song, P., Fan, Y., Kalbfleisch, J.D.: Maximization by parts in likelihood inference. Technical Report, Vanderbilt University. http://econpapers.repec.org/paper/vanwpaper/0319.htm (2003)

Zolotarev, V.: Modern Theory of Summation of Random Variables. VSP, Utrecht (1997)