New Copula Models in Quantitative Finance

with Applications to Risk Management and Option Pricing

Alan Smillie

Centre for Quantitative Finance
Tanaka Business School
Imperial College of Science, Technology and Medicine
University of London

May 22, 2007

Submitted for the degree of Doctor of Philosophy

Abstract

The study of multivariate distributions with given marginals - or copulas - has been the subject of much recent research in quantitative finance. We review the principal approaches in the literature, and asses their practical utility. In particular we seek to re-examine the measurement of dependence in the extremes, or tail dependence.

We introduce the Gaussian Mixture copula, which we believe offers an improvement over the existing parametric models, and we explore the properties of this new model and show how it can be implemented in practice. We also consider the non-parametric modelling of dependence, extend some of the underlying results and propose innovative methods for the implementation of such models. Finally we present two extended examples of how the new models might be applied to problems in quantitative finance, and argue that they give better results than the existing approaches.

Acknowledgements

Many thanks to everyone who helped with my thesis, in particular my supervisor Prof. Nicos Christofides of the Centre for Quantitative Finance at Imperial College, my sponsor Dr. Eduardo Epperlein of Citigroup Risk Architecture, and my colleagues at the Tanaka Business School and Citigroup Market Risk Analytics.

I am also very grateful to my friends and family for all their support, but especially to my parents Gordon and Christine for always encouraging me to study and to Joanna for her gentle reminders that this thesis wouldn't write itself.

Contents

1 Motivation & Outline **1**

2 Theory of Copulas **4**

2.1 Definitions & Basic Properties 4

2.2 Dependence Measures 8

2.2.1 Linear Correlation 9

2.2.2 Kendall's Tau 10

2.2.3 Spearman's Rho 11

2.2.4 Tail Dependence 12

2.3 Implementation 16

2.3.1 Calibration 16

2.3.2 Simulation 18

3 Parametric Copula Models **20**

3.1 Existing Parametric Models 20

3.1.1 Gaussian Copula 20

3.1.2 Student's t copula and extensions 23

3.1.3 Archimedean Copulas 26

3.1.4 Mixture Copulas 30

3.2 Gaussian Mixture Copula 31

3.2.1 Definitions 31

3.2.2 Dependence Measures 34

3.2.3 Implementation 42

4 Non-Parametric Copula Models **46**

4.1 Empirical Sub-Copula 46

4.2 Partitions of Unity 48

4.2.1 Checkerboard Approximation 51
4.2.2 Bernstein Approximation 52
4.3 Kernel Estimators 56
4.4 Implementation 60
4.4.1 Calibration 60
4.4.2 Simulation 63

5 Applications of Copulas in Quantitative Finance **70**
5.1 Portfolio Risk Management 70
5.1.1 Risk Measures 70
5.1.2 Modelling Market Risk 71
5.1.3 Portfolio Experiments 74
5.2 Pricing Multi-asset Derivatives 78
5.2.1 Pricing Theory 78
5.2.2 Option Pricing Formulae 80
5.2.3 Pricing with Copulas 84

6 Summary **92**

A Probability Theory **94**
A.1 Elementary Probability Theory 94
A.2 Random Variables & Distributions 95
A.3 Elliptical Distributions 98
A.4 Random Processes 99

B Summary Statistics for Sample Data **100**

Chapter 1

Motivation & Outline

How astonishing and admirable are the ways of Providence, what thoughts come to our mind when observing the marvellous order which presides over the most minute details of the most hidden events. Why, the changes in stock market prices are subject to fixed mathematical laws! Events produced by the passing fancy of men, the most unpredictable shocks of the political world, of clever financial schemes, the outcome of a vast number of unrelated events, all this combines and randomness becomes a word without meaning.

Jules Regnault, translated from the French
Calcul des Chances et Philosophie de la Bourse
Paris, 1863

Ever since Regnault's day researchers have striven to use the laws of mathematics to describe the behaviour of the market. The most successful descriptions have been those based around the normal or log-normal distribution: cornerstones of modern finance like Markowitz-Sharpe portfolio theory and the Black-Scholes-Merton framework for option pricing have normality at their core. But, even as quantitative modelling becomes ever more accepted by market practioners, the limitations of some of its basic assumptions have become increasingly evident. Whilst researchers have studied alternatives to normality for the one-dimensional problem for many years (see, for example, [Man 63] and [Fam 65]), it is only recently that the modelling of dependence between financial assets has been extended beyond the standard, covariance-based model ([Emb 01] and [Che 04]).

The canonical measure of statistical dependence is the *copula* function which, together with a set of marginal distributions, completely determines the multivariate dis-

tribution of a set of random variables. It is therefore natural to use copulas as a tool in the quantitative analysis of the relationship between financial assets, and there has been much recent interest in this field. Research has focused on the development of models capable of capturing certain empirically observed stylised facts which are not captured by the standard model: asset returns tend to make large co-movements more frequently than is implied by the Gaussian model; and such comovements tend to be asymmetric - that is they occur in one direction more than the other. Such observations can intuitively be thought of as demonstrating that market crashes occur with greater frequency and greater severity than the standard model leads us to expect. Blindly applying the Gaussian model can thus lead to serious problems: underestimation of the risk in a supposedly diversified portfolio or mispricing of so called 'rainbow options' written on multiple underlyings, for example.

Formulating a model capable of capturing such behaviour is, however, only a part of the solution. The standard model became and largely remains the standard for good reason: it is simple to understand and easy to implement. In attempting to move beyond the Gaussian world the researcher is constantly forced to balance the intellectual appeal of a 'perfect' fit to empirical data with the practioner's priorities of simplicity and robustness. In what follows we have tried to develop methods which achieve such a compromise.

The thesis divides into four main parts: we begin by giving a formal mathematical definition of copula functions and their use in multivariate statistics, together with some related concepts. A collection of measures of statistical association from the literature is described and we introduce a new measure which we consider to be useful in our applications. We then describe some of the general approaches to calibrating copulas to empirical data and to simulating from some chosen model: two key practical issues which will recur throughout the thesis.

The next chapter outlines the copula models that can be applied to financial modelling. The principal existing models - Gaussian, Student's t and Archimedean - are reviewed, together with some proposed extensions such as the Modified Student's t and the Mixture copula. Next we describe the first of our main original contributions: the Gaussian Mixture copula. We study the properties of our new model, derive expressions for the dependence measures and show how it could be implemented in practice.

A fundamentally different approach is to use a non-parametric copula, where the dependency structure is driven directly by some empirical data, rather than some parametric model chosen *a priori*. We give a rigorous definition of the ideas that form the

basis of this approach, and go on to describe some specific methods of constructing non-parametric copulas. The non-parametric method has received relatively little attention in the applied literature, so we pay particular attention to the implementation of these models in practice.

In the last of the main chapters we consider the application of copula modelling to two of the classic problems in quantitative finance: portfolio risk management and option pricing. Each topic is briefly introduced, and the implications of non-normal dependence on each are analysed in more detail. We present an extended worked example on each topic, demonstrating how the new models developed in the previous chapter can be applied and what impact they have.

Chapter 2

Theory of Copulas

2.1 Definitions & Basic Properties

We begin by defining formally the copula function in d-dimensions [Nel 99]:

Definition 1 *A* d-dimensional copula *is a function* $C : [0,1]^d \to [0,1]$ *such that*

1. *C is grounded; i.e. for all* $\mathbf{u}$ *in* $[0,1]^d$, $C(u_1, \ldots, u_d) = 0$ *if at least one component* $u_k = 0$.

2. *C has uniform marginals; i.e. for all* u_k *in* $[0,1]$, $C(1, \ldots, 1, u_k, 1, \ldots, 1) = u_k$.

3. *C is d-increasing; i.e for any* $\mathbf{a}$ *and* $\mathbf{b}$ *in* $[0,1]^d$ *such that* $\mathbf{a} \leq \mathbf{b}$,

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C(u_{1i_1}, \ldots, u_{di_d}) \geq 0,$$

where $u_{j1} = a_j$ *and* $u_{j2} = b_j$ *for* $j = 1, \ldots, d$.

Observe that the first and third conditions in the above definition suffice to ensure that a copula satisfies the requirements to be a d-dimensional probability distribution function[1] on $[0,1]^d$, while the second condition ensures that the marginal distributions are uniform.

In what follows will also use the concept of a sub-copula

Definition 2 *A* d-dimensional sub-copula C' *is a grounded, d-increasing function with uniform marginals having* $\operatorname{Dom} C' = S_1 \times S_2 \times \cdots \times S_d$ *where each* S_j *is a subset of* $[0,1]$ *containing* 0 *and* 1.

[1] We assume that the reader is familiar with standard concepts from probability theory such as probability distribution and density functions. A brief review of these ideas is given in Appendix A

Clearly a copula is just a sub-copula having domain $[0,1]^d$.

We note the following definitions and theorems which will be useful later on, for proofs and further details see [Joe 97] or [Nel 99].

Theorem 1 *Let C' be a d-dimensional sub-copula. Then for all* $\mathbf{u}$ *and* $\mathbf{v}$ *in* $\operatorname{Dom} C'$,

$$|C'(\mathbf{u}) - C'(\mathbf{v})| \leq \sum_{k=1}^{d} |v_k - u_k| \tag{2.1}$$

i.e. C' is uniformly continuous on its domain.

Proof See [Nel 99], §2.10.

The importance of copulas in multivariate statistical modelling derives from the following theorem and its corollary:

Theorem 2 Sklar's Theorem *Let F be a d-dimensional distribution function with marginal distributions $F_1, F_2, \ldots, F_d$. Then there exists a copula C such that, for all* $\mathbf{x} \in (-\infty, \infty)^d$,

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)).$$

Furthermore, if $F_1, F_2, \ldots, F_d$ are continuous, then C is unique; otherwise C is unique on $\operatorname{Ran} F_1 \times \ldots \times \operatorname{Ran} F_d$.

Proof See [Skl 59] for the proof for the two-dimensional case, and [Skl 96] for the d-dimensional version.

Corollary 3 *Let F be a d-dimensional distribution function with continuous marginals $F_1, F_2, \ldots, F_d$, and let C be a d-dimensional copula. Then, for any* $\mathbf{u} \in [0,1]^d$,

$$C(u_1, u_2, \ldots, u_d) = F(F_1^{(-1)}(u_1), F_2^{(-1)}(u_2), \ldots, F_d^{(-1)}(u_d)) \tag{2.2}$$

where $F_1^{(-1)}, F_2^{(-1)}, \ldots, F_d^{(-1)}$ are the quasi-inverses of the marginal distributions.

Thus we can see that given *any* combination of marginal distributions and *any* choice of copula we can generate a multivariate probability distribution. It is this property of the copula, which allows us to separate the modelling of the univariate marginals from the modelling of the dependency, that is most useful from a financial modelling perspective. Also notice how the copula can be regarded as the distribution function of the quantiles $U_i = F_i(X_i)$, since

$$C(u_1, \ldots, u_d) = F(F_1^{(-1)}(u_1), \ldots, F_d^{(-1)}(u_d))$$

$$= P(X_1 \leq F_1^{(-1)}(u_1), \ldots, X_d \leq F_d^{(-1)}(u_d))$$
$$= P(F_1(X_1) \leq F_1(F_1^{(-1)}(u_1)), \ldots, F_d(X_d) \leq F_d(F_d^{(-1)}(u_d)))$$
$$= P(U_1 \leq u_1, \ldots, U_d \leq u_d).$$

We can loosely interpret the copula as telling us the probability that some collection of random variables $X_1, \ldots, X_d$ take 'low' (i.e. less than $u_1, \ldots, u_d$) values simultaneously.

Next, we make a definition which will show how a copula can be viewed as density function rather than a distribution function. We have that f, the d-dimensional probability density function of of F, is

$$f(x_1, \ldots, x_d) = \frac{\partial^d F(x_1, \ldots, x_d)}{\partial x_1 \ldots \partial x_d} = \frac{\partial^d C(F_1(x_1), \ldots, F_d(x_d))}{\partial x_1 \ldots \partial x_d}$$
$$= \frac{\partial^d C(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d} \prod_{i=1}^{d} f_i(x_i) \tag{2.3}$$

where $u_i = F_i(x_i)$ and f_i is the pdf of the ith marginal. Hence we define the copula density as

Definition 3 *The* copula density *of a d-dimensional copula C is the function $c : [0,1]^d \to [0, \infty)$ such that*

$$c(u_1, \ldots, u_d) \equiv \frac{\partial^d C(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d} = \frac{f(x_1, \ldots, x_d)}{\prod_{i=1}^{d} f_i(x_i)} \tag{2.4}$$

Recall that the continuous random variables $X_1, \ldots, X_d$ are called *independent* if and only if have distribution function

$$F^{\perp}(X_1, \ldots, X_d) \equiv \prod_{i=1}^{d} F_i(X_i)$$

and density function

$$f^{\perp}(X_1, \ldots, X_d) \equiv \prod_{i=1}^{d} f_i(X_i).$$

Applying (2.2) and (2.3) to the aobove definitions we can see that a necessary and sufficient condition for the independence of $X_1, \ldots, X_d$ is that they have the copula

$$C^{\perp}(u_1, \ldots, u_d) \equiv \prod_{i=1}^{d} u_i,$$

with copula density

$$c^{\perp}(u_1, \ldots, u_d) = 1.$$

$C^{\perp}$ is referred to as the *product copula.*

We can construct upper and lower limits for copulas in the form of the Frechet bounds.

Theorem 4 Frechet bounds *Let C be a d-dimensional copula. For all* $\mathbf{u} \in [0,1]^d$

$$\max(u_1 + \cdots + u_d - 1, 0) < C(u_1, \ldots, u_d) \leq \min(u_1, \ldots, u_d).$$

Proof See [Nel 99], §2.10.

We call

$$C^-(u_1, \ldots, u_d) \equiv \max(u_1 + \cdots + u_d - 1, 0)$$

the *lower Frechet bound* and

$$C^+(u_1, \ldots, u_d) \equiv \min(u_1, \ldots, u_d)$$

the *upper Frechet bound.* We remark that the density of C^+ is zero everywhere away from the line $u_1 = u_2 = \cdots = u_d$, while the density of C^- is zero away from the $d-1$ dimensional hypersurface defined by $u_1 + \cdots + u_d = 1$. Also note that while C^- forms a lower bound for the set of copula functions, it is not itself a copula for $d > 2$.

It is also possible to define the copula analogue of a multivariate survival function (see Appendix A.2 for the definition), which we do via the survival version of Sklar's theorem

Definition 4 *Given some multivariate survival function $\bar{F}$, the* survival copula *is the function $\tilde{C} : [0,1]^d \to [0,1]$ such that*

$$\bar{F}(\mathbf{x}) = \tilde{C}(\bar{F}_1(x_1), \ldots, \bar{F}_d(x_d))$$

Notice how, just as the copula tells us about the probability two random variables are below some particular quantile, the survival copula tells us something about the probability that two random variables are jointly greater than some values $u_1, \ldots, u_d$

$$\tilde{C}(1 - u_1, \ldots, 1 - u_d) = P(U_1 > u_1, \ldots, U_d > u_d).$$

For the bivariate case we can express $\tilde{C}$ in terms of C as

$$\tilde{C}(1 - u_1, 1 - u_2) = 1 - u_1 - u_2 + C(u_1, u_2).$$

A multivariate extension of the above expression is available (see [Geo 01]), but the expression is complex and omit it here.

The last basic property we mention concerns the symmetry of bivariate distributions and copulas.

Definition 5 *A bivariate random vector* (X_1, X_2) *is said to be* radially symmetric about (a, b) *if and only if the joint distribution function of* $(X - a, Y - b)$ *is the same as the joint distribution function of* $(a - X, b - Y)$. *This is equivalent to requiring that*

$$F(a + x_1, b + x_2) = \bar{F}(a - x_1, b - x_2) \qquad \text{for all } x_1, x_2$$

Theorem 5 *Suppose we have a bivariate random vector* (X_1, X_2), *and that* X_1 *and* X_2 *are symmetric about some points* a *and* b *respectively. Then* (X_1, X_2) *is radially symmetric about* (a, b), *i.e.* $F(a + x_1, b + x_2) = \tilde{F}(a - x_1, b - x_2)$, *if and only if*

$$C(u_1, u_2) = \tilde{C}(u_1, u_2). \tag{2.5}$$

A copula function that satisfies 2.5 is said to be radially symmetric.

Proof See [Nel 99], §2.7.

When copulas are used to model the dependence between equity prices, radial symmetry has an important economic interpretation, since it implies that

$$P(U_1 \leq u, U_2 \leq u) = P(U_1 > 1 - u, U_2 > 1 - u),$$

i.e. that the probability of a joint upward move (i.e. a rally) is the same as the probability of a joint downward move (a crash).

Before looking further at copulas and their applications in finance, we introduce the data we shall use to illustrate some these ideas and to give examples of how our innovative techniques can be applied. We use daily returns data from a collection of Asian equity indexes[2] between 06-Jan-1982 and 01-Jun-2006. The dataset consists of Hong Knog's Hang Seng Price Index (HK), Japan's Nikkei 225 Stock Price Index (JP), Korea's Kospi Index (KO), the Singapore Straits Times Price Index (SG), Thailand's Bangkok Book Club Index (TH) and the Taiwan Stock Exchange Corporation Index (TW). A range of summary statistics for this data can be found in Appendix B, we shall use the SG-TH pair to give examples for the bivariate case. Figure 2.1 shows a scatter plot of this data.

2.2 Dependence Measures

In the study of copulas it is very useful to have some measure of the dependence between random variables. It is generally most convenient to do this pairwise, and we now introduce some of the most commonly used measures of pairwise dependence below.

[2]Sourced from Reuters on 03-Jun-2006

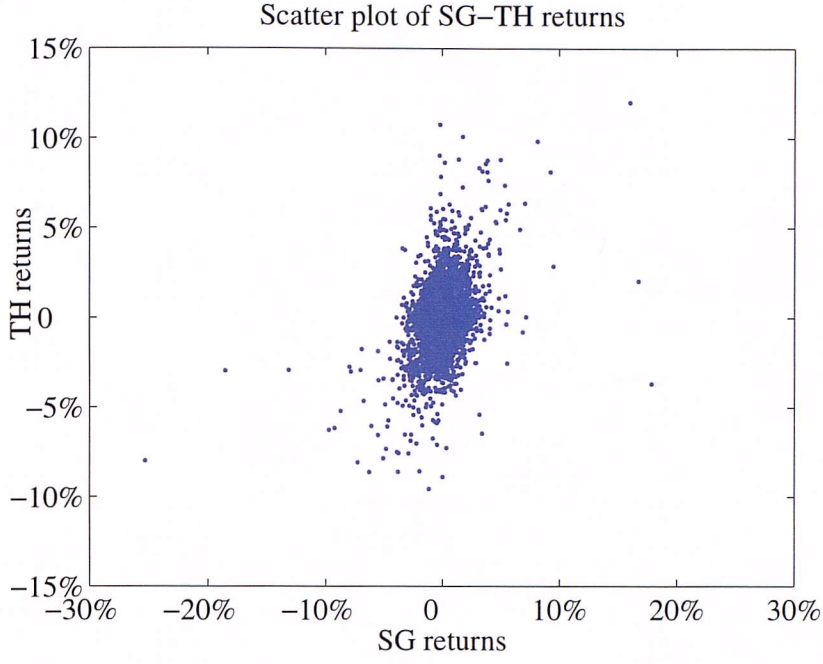


Figure 2.1: SG-TH returns

2.2.1 Linear Correlation

The best known measure of dependence is the linear, or Pearson's, correlation:

Definition 6 *Given a pair of random variables X_1 and X_2, each having finite variance, the* linear correlation *is*

$$\rho(X_1, X_2) = \frac{\mathrm{cov}(X_1, X_2)}{\sqrt{\mathrm{var}(X_1)\mathrm{var}(X_2)}} \tag{2.6}$$

where $\mathrm{var}(\cdot)$ *and* $\mathrm{cov}(\cdot\,,\cdot)$ *denote variance and covariance respectively.*

Linear correlation is widely used since it is the natural measure of dependence for the Gaussian distribution (and more generally for the class of elliptical distributions, see §3.1) and since it is straightforward to compute from empirical data: given a bivariate sample $\{\bar{x}_1^{(t)}, \bar{x}_2^{(t)}\}_{t=1}^T$ we have that the sample correlation $\bar{\rho}$ is

$$\bar{\rho}(\bar{x}_1, \bar{x}_2) = \frac{\sum_{t=1}^T (\bar{x}_1^{(t)} - \bar{\mu}_1)(\bar{x}_2^{(t)} - \bar{\mu}_2)}{\sqrt{\sum_{t=1}^T (\bar{x}_1^{(t)} - \bar{\mu}_1)^2}\sqrt{\sum_{t=1}^T (\bar{x}_1^{(t)} - \bar{\mu}_1)^2}},$$

where $\bar{\mu}$ denotes sample mean. For the sample data set we have

$$\bar{\rho}(\mathrm{SG}, \mathrm{TH}) = 0.37$$

Linear correlation is not in general the most useful measure of dependence, in large part because it is not invariant under increasing transformations, i.e. there exist functions $f(X)$ such that $X > Y \Rightarrow f(X) > f(Y)$ but

$$\rho(f(X_1), f(X_2)) \neq \rho(X_1, X_2).$$

This can lead to the situations where a pair of random variables X_1 and X_2 are perfectly co-monotonic , but $\rho(X_1, X_2) \neq 1$. For example, consider the log-normally distributed random variables defined by

$$X_1 = e^Z \quad \text{and} \quad X_2 = e^{\sigma Z},$$

where Z has a standard normal distribution. Clearly X_1 and X_2 exhibit perfect dependence, but it can be shown (see [Emb 01]) that

$$\lim_{\sigma \to \infty} \rho(X_1, X_2) = 0$$

This dependence on the marginals is made clear by expressing ρ in terms of a copula

$$\rho(C, F_1, F_2) = \frac{1}{\sigma_1 \sigma_2} \int_{\cdot} \int_{[0,1]^2} [C(u_1, u_2) - u_1 u_2] \, dF_1^{-1}(u_1) dF_2^{-1}(u_2).$$

Evidently, linear correlation depends on the form of the marginals F_1 and F_2 as well as the copula C. We also have the problem that linear correlation is not even defined for random variables with infinite variance, which creates difficulties when using heavy tailed distributions to construct a model.

2.2.2 Kendall's Tau

A more robust measure of dependence is concordance – a pair of observations (x_1, x_2) and $(\acute{x_1}, \acute{x_2})$ from the random vector (X_1, X_2) are said to be concordant if $(x_1 - \acute{x_1})(x_2 - \acute{x_2}) > 0$ and discordant if $(x_1 - \acute{x_1})(x_2 - \acute{x_2}) < 0$. We can measure the level of concordance between random vectors by Kendall's tau:

Definition 7 *Given a pair of random variables X_1 and X_2,* Kendall's tau *is*

$$\tau(X_1, X_2) \equiv \mathbf{P}[(X_1 - \acute{X_1})(X_2 - \acute{X_2}) > 0] - \mathbf{P}[(X_1 - \acute{X_1})(X_2 - \acute{X_2}) < 0],$$

where $(\acute{X_1}, \acute{X_2})$ is an independent sample from the distribution of (X_1, X_2).

Thus we can see that Kendall's tau is the probability of concordance minus the probability of discordance. Given that (X_1, X_2) are joined by a copula C, we have [Nel 99] that

$$\tau(C) = 4 \int_{\cdot} \int_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 1.$$

So, unlike linear correlation, Kendall's tau for a pair of random variables is completely determined by their copula. This result will enable us to express the τ of a range of parametric copulas as a function their parameters, which will be useful later on when we need calibrate non-elliptical copulas to market data.

Estimating τ from empirical data is straightforward: using the signum function

$$\text{sgn}(x) \equiv \begin{cases} -1 & \text{if} \quad x < 0 \\ 0 & \text{if} \quad x = 0 \\ 1 & \text{if} \quad x > 0, \end{cases}$$

we can write that the sample Kendall's tau is

$$\bar{\tau}(\bar{x}_1, \bar{x}_2) = \frac{2}{T(T-1)} \sum_{s=1}^{T} \sum_{t>s} \text{sgn}\left((\bar{x}_1^{(s)} - \bar{x}_1^{(t)})(\bar{x}_2^{(s)} - \bar{x}_2^{(t)})\right).$$

Note that this algorithm is $O(T^2)$, unlike that for computing linear correlation which is $O(T)$, which may cause some computational problems when working with long time series. Using this method on the sample data we have

$$\bar{\tau}(\text{SG}, \text{TH}) = 0.21$$

By analogy with the familiar correlation matrix, we define the *Kendall's tau matrix* for a d-dimensional distribution to be $T \in [-1, 1]^{d \times d}$ such that

$$T_{ij} = \tau(X_i, X_j) \quad \text{for} \quad i, j = 1, \ldots, d.$$

2.2.3 Spearman's Rho

Another robust measure of dependence, closely related to Kendall's tau, is Spearman's rho.

Definition 8 *Gievn a pair of random variables X_1 and X_2,* Spearman's rho *is*

$$\rho_s(X_1, X_2) \equiv 3\left(\mathbf{P}[(X_1 - \acute{X}_1)(X_2 - \grave{X}_2) > 0] - \mathbf{P}[(X_1 - \acute{X}_1)(X_2 - \grave{X}_2) < 0]\right),$$

where $(\acute{X}_1, \acute{X}_2)$ and $(\grave{X}_1, \grave{X}_2)$ are independent samples from the distribution of (X_1, X_2).

We can express ρ_s in copula terms as

$$\rho_s(C) = 12 \int \int_{[0,1]^2} [C(u_1, u_2) - u_1 u_2] du_1 du_2,$$

which will again enable us to compute ρ_s for a range of parametric copulas. Clearly ρ_s, like τ, does not depend on the form the marginal distributions.

It can be shown (see, for example, [Nel 99] pg 137) that Spearman's rho can be regarded as measuring the rank correlation, i.e.

$$\rho_s(X_1, X_2) = \rho(U_1, U_2) = 12\text{cov}(U_1, U_2),$$

where $U_i = F_i(X_i)$. Thus, to estimate $\bar{\rho}_s$ form an empirical sample we need only compute the order statistics $\bar{u}_i$ using (2.13), then estimate the linear correlation as usual, i.e.

$$\bar{\rho}_s(\bar{x}_1, \bar{x}_2) = \frac{\sum_{t=1}^{T}(\bar{u}_1^{(t)}\bar{u}_2^{(t)})}{\sqrt{\sum_{t=1}^{T}(\bar{u}_1^{(t)})^2}\sqrt{\sum_{t=1}^{T}(\bar{u}_2^{(t)})^2}} = 12\frac{\sum_{t=1}^{T}(\bar{u}_1^{(t)}\bar{u}_2^{(t)})}{T(T^2-1)}.$$

This algorithm is $O(T)$ and the construction of the rank statistics $\bar{u}$ is $O(T\log T)$, so for long time series it may be preferable to compute $\bar{\rho}_s$ in favour of $\bar{\tau}$. The sample data has

$$\bar{\rho}_s(\text{SG}, \text{TH}) = 0.30$$

For multivariate distributions we can construct a Spearman's rho matrix $S \in [-1, 1]^{d\times d}$ as

$$S_{ij} = \rho_s(X_i, X_j) \quad \text{for} \quad i, j = 1, \ldots, d.$$

2.2.4 Tail Dependence

A rather different measure of dependence that has received attention in the financial literature is the concept of tail dependence.

Definition 9 *Given random variables X_1 and X_2 having marginals F_1 and F_2, the* upper tail dependence coefficient *and* lower tail dependence coefficient *are respectively*

$$\begin{aligned}\lambda_U(X_1, X_2) &\equiv \lim_{u\to 1^-} \mathbf{P}[X_2 > F_2^{-1}(u) | X_1 > F_1^{-1}(u)] \\ &= \lim_{u\to 1^-} \mathbf{P}[X_1 > F_1^{-1}(u) | X_2 > F_2^{-1}(u)]\end{aligned}$$

and

$$\begin{aligned}\lambda_L(X_1, X_2) &\equiv \lim_{u\to 0^+} \mathbf{P}[X_2 \leq F_2^{-1}(u) | X_1 \leq F_1^{-1}(u)] \\ &= \lim_{u\to 0^+} \mathbf{P}[X_1 \leq F_1^{-1}(u) | X_2 \leq F_2^{-1}(u)]\end{aligned}$$

A copula is said to be upper tail dependent *if $\lambda_U \neq 0$ and* upper tail independent *if $\lambda_U = 0$, and* lower tail dependent *if $\lambda_L \neq 0$ and* lower tail independent *if $\lambda_L = 0$.*

Intuitively, we can interpret λ_U and λ_L as a measuring of the likelihood that one variable takes an 'extreme value', given that the other does. Note, however, that 'upper tail independence' is *not* equivalent to 'independence in the tail', since $\lambda_U = 0$ does *not* imply that

$$\lim_{x_1,x_2\to\infty} \frac{\mathbf{P}[x_1 \leq X_1, x_2 \leq X_2]}{\mathbf{P}[x_1 \leq X_1]\mathbf{P}[x_2 \leq X_2]} = 1,$$

and similarly for the lower tail.

[Joe 97] shows that the above expressions can be written in terms of the copula function C as

$$\lambda_U(C) = \lim_{u\to 1^-} \frac{1 - 2u + C(u,u)}{1-u} \quad \text{and} \quad \lambda_L(C) = \lim_{u\to 0^+} \frac{C(u,u)}{u}. \tag{2.7}$$

Combing (2.7) and (2.5), we can see that all radially symmetric copulas have $\lambda_L = \lambda_U$.

In contrast to the dependence measures we have considered previously, measuring the tail dependence coefficient of empirical data is highly problematic. This is largely because λ_U and λ_L are defined in the limit of $u \to 1^-$ and $u \to 0^+$. This problem is discussed in detail by [Fra 05], who propose the following non-parametric estimators (in our notation) for λ_U:

$$\hat{\lambda}_U(u) = \frac{1}{T(1-u)} \sum_{t=1}^{T} \mathbf{1}_{[\bar{u}_1^{(t)} > u \,\wedge\, \bar{u}_2^{(t)} > u]} \tag{2.8}$$

and

$$\bar{\lambda}_U(k) = 2 - \frac{\log\left(\frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{[\bar{r}_1^{(t)} \leq u \,\wedge\, \bar{r}_2^{(t)} \leq u]}\right)}{\log(1-u)}, \tag{2.9}$$

and

$$\tilde{\lambda}_U = 2 - 2\exp\left(\frac{1}{T}\sum_{t=1}^{T} \log\left\{\sqrt{\log\frac{1}{r_1^{(t)}}\log\frac{1}{r_2^{(t)}}} \Big/ \log\frac{1}{\max\left(r_1^{(t)}, r_2^{(t)}\right)^2}\right\}\right), \tag{2.10}$$

with analogous expressions for λ_L. For estimators (2.8) and (2.9) we must make a compromise in choosing $u < 1$ to minimise the variance and bias of the estimator, which makes it very difficult to determine with confidence whether an empirical data set is tail dependent. Estimator (2.10) avoids this problem, but only at the cost of assuming *a priori* that the true copula does exhibit some tail dependence, and then measuring how much. Unsurprisingly, this creates large bias (see [Fra 05]) when the copula from which the sample is drawn has $\lambda = 0$.

In §3.2 we shall present an examples of copulas that exhibit perfect tail dependence, but for which the above estimators would yield $\bar{\lambda_U} \approx 0$, and vice versa. Such fundamental uncertainty in estimating λ_L and λ_U empirically leads us to question their

usefulness: what do we gain by attempting to measure, and calibrate to, an intrinsically hard to measure quantity? It is our opinion that, while λ_L and λ_U as defined above may be of some interest in classifying parametric models, the fundamental difficulties encountered in trying to measure them empirically render them of little interest to the applied researcher, and that the emphasis given to them in much of the 'copulas for finance' literature is misplaced.

When one attempts to measure tail dependence the estimators 2.8 and 2.9 one is in fact measuring $C(\frac{k}{T}, \frac{k}{T})/\frac{k}{T}$ for some small *but non-zero* value $\frac{k}{T}$. We therefor suggest that instead of measuring $\bar{C}(u,u)/u$ for some small u, assuming that this is equivalent to $\bar{\lambda_L}$, and comparing it with λ_L from some parametric model; the researcher should simply measure $\bar{C}(u,u)$ for u small, and compare this directly with $C(u,u)$ of the parametric model of interest. A similar procedure for upper tail dependence would involve constructing an empirical estimate of $\tilde{C}(u,u)$ and comparing this with the model value. To this end, we propose a dependence[3] measure which can act as a robust alternative to upper and lower tail dependence.

Definition 10 *Given random vectors X_1 and X_2 the* lower quantile dependence *is the function $\nu_L : [0,1] \to [0,1]$ such that*

$$\begin{aligned} \nu_L(u; X_1, X_2) &= \mathbf{P}[X_2 \leq F_2^{-1}(u) | X_1 \leq F_1^{-1}(u)] \\ &= \mathbf{P}[X_1 \leq F_1^{-1}(u) | X_2 \leq F_2^{-1}(u)] \end{aligned} \tag{2.11}$$

and the upper quantile dependence *is the function $\nu_U : [0,1] \to [0,1]$ such that*

$$\begin{aligned} \nu_U(u; X_1, X_2) &= \mathbf{P}[X_2 > F_2^{-1}(u) | X_1 > F_1^{-1}(u)] \\ &= \mathbf{P}[X_1 > F_1^{-1}(u) | X_2 > F_2^{-1}(u)] \end{aligned} \tag{2.12}$$

These can be expressed very simply in copula terms as

$$\nu_L(u) = \frac{C(u,u)}{u} \qquad \text{and} \qquad \nu_U(u) = \frac{1 - 2u + C(u,u)}{1-u} = \frac{\tilde{C}(u,u)}{1-u}.$$

Clearly, radial symmetry implies that $\nu_L(u) = \nu_U(1-u)$ for all $u \in (0,1)$.

We can easily construct empirical estimators for $\lambda_L(u)$ and $\lambda_U(u)$ as

$$\bar{\nu}_L(u) = \frac{1}{Tu} \sum_{t=1}^{T} \mathbf{1}_{[\bar{u}_1^{(t)} \leq u \wedge \bar{u}_2^{(t)} \leq u]}$$

[3] A similar estimator is mentioned by [Col 99] as a way of measuring extremes in meteorological data. These authors focus on measuring a modified form of (2.11) and (2.12), $\bar{\chi}(u) = \frac{2\log(1-u)}{\log \bar{C}(1-u,1-u)}$.

and

$$\bar{\nu}_U(u) = \frac{1}{T(1-u)} \sum_{t=1}^{T} \mathbf{1}_{[\bar{u}_1^{(t)} > u \wedge \bar{u}_2^{(t)} > u]}.$$

Using these estimators we have plotted $\bar{\nu}_L$ and $\bar{\nu}_U$ for the sample data in Figure 2.2: $\bar{\nu}_L(u)$ is shown in the interval $u \in (0, 0.5]$, and $\bar{\nu}_U(u)$ in the region $u \in [0.5, 1)$. Notice how the sample data exhibit somewhat greater dependence in the lower than in the upper tail (i.e. $\nu_L(u) > \nu_U(1-u)$ as $u \to 0$). This behaviour has been observed consistently for time series of equity returns (e.g. [Lon 01] and [Ang 02]) and will be important when we consider the problems of portfolio risk and the pricing of multivariate contingent claims.

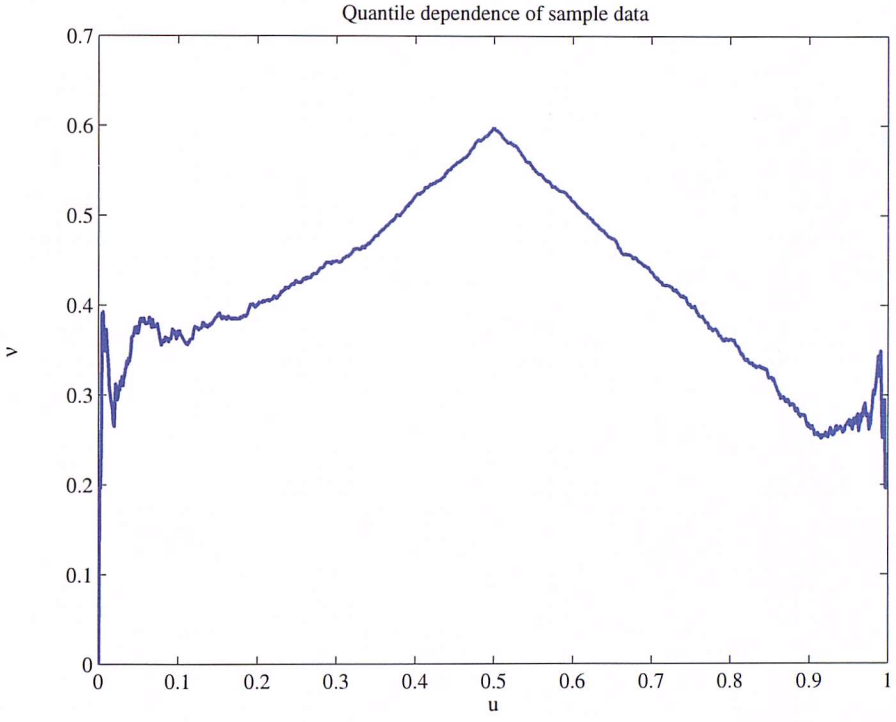


Figure 2.2: Quantile Dependence of SG - TH returns

For multivariate distributions we can construct tail dependence matrices $L_L \in \mathbf{R}^{d\times d}$ and $L_U \in \mathbf{R}^{d\times d}$ and quadrant dependence matrices $N_L \in \mathbf{R}^{d\times d}$ and $N_U \in \mathbf{R}^{d\times d}$ in the obvious fashion

$$L_{L_{ij}} = \lambda_L(X_i, X_j) \quad \text{and} \quad L_{U_{ij}} = \lambda_U(X_i, X_j) \quad \text{for} \quad i, j = 1, \ldots, d,$$

$$N_{L_{ij}} = \nu_L(X_i, X_j) \quad \text{and} \quad N_{U_{ij}} = \nu_U(X_i, X_j) \quad \text{for} \quad i, j = 1, \ldots, d.$$

2.3 Implementation

The two main challenges in implementing copula techniques lie in calibrating our chosen model to market data, and - if we wish to use Monte Carlo techniques - generating samples from the chosen copula.

2.3.1 Calibration

There are in principle three ways to calibrate a parametric copula.

One Step Maximum Likelihood Suppose we have a set of multivariate empirical data to which we wish to calibrate some multivariate model having copula C and marginals $F_1, \ldots, F_d$. Denote the parameters of the copula by θ_c, the parameters of the ith marginal by θ_i and the entire parameter set as $\theta = \{\theta_c, \theta_1, \ldots, \theta_d\}$, where θ takes values in Θ. Let us construct a log likelihood function l of the form

$$l(\theta) = \sum_{t=1}^{T} \ln c(F_1(\bar{x}_1^{(t)}), \ldots, F_d(\bar{x}_d^{(t)})) + \sum_{t=1}^{T} \sum_{i=1}^{d} \ln f_i(\bar{x}_i^{(t)}).$$

Using the familiar technique of maximum likelihood estimation we can construct $\bar{\theta}$, an empirical estimate of θ, as

$$\bar{\theta} = \max_{\theta \in \Theta} l(\theta).$$

The apparent simplicity of this approach is attractive, but the large number of parameters to be estimated simultaneously makes this approach slow and unreliable in practice

Two Step Maximum Likelihood The one step MLE described above can be difficult to implement because of the potentially high number of parameters involved, a problem that becomes especially severe in high dimension. Instead we can exploit the key property of copula to separate out the calibration of the marginals from the calibration of the copula. First, we calibrate the marginals by maximising

$$l_i(\theta_i) = \sum_{t=1}^{T} \ln f_i(\bar{x}_i^{(t)})$$

for each marginal to find $\{\bar{\theta}_1, \ldots, \bar{\theta}_d\}$. We then calibrate the copula by maximising

$$l_c(\theta_c; \bar{\theta}_1, \ldots, \bar{\theta}_d) = \sum_{t=1}^{T} \ln c(F_1(\bar{x}_1^{(t)}; \bar{\theta}_d), \ldots, F_d(\bar{x}_d^{(t)}; \bar{\theta}_d)).$$

In fact we can extend this approach and avoid calibrating the marginals completely. Denoting a d-dimensional empirical data set set of length T by $\{\bar{x}_1^{(t)}, \ldots, \bar{x}_d^{(t)}\}_{t=1}^T$, we make this uniform on $[0,1]$ we apply the *empirical marginal transform* defined as

$$\bar{F}(x; \bar{x}) = \frac{1}{T+1} \sum_{t=1}^{T} \mathbf{1}_{[\bar{x}^{(t)} \leq x]}, \tag{2.13}$$

where $\mathbf{1}_{[\cdot]}$ denotes the indicator function of $[\cdot]$. Then we have that the transformed empirical sample is given by

$$\{\bar{u}_1^{(t)}, \ldots, \bar{u}_d^{(t)}\}_{t=1}^T = \{\bar{F}(\bar{x}_1^{(t)}), \ldots, \bar{F}(\bar{x}_d^{(t)})\}_{t=1}^T.$$

Notice that this is closely linked to the rank statistic of the data set. Letting $\bar{r}_i^{(t)}$ denote the rank of $\bar{x}_i^{(t)}$,

$$\bar{u}_i^{(t)} = \frac{\bar{r}_i^{(t)}}{T+1}.$$

This is useful since there exists an $O(T \ln T)$ algorithm for the construction of rank statistics (see [Pre 92] pp. 338 - 341), which we can easily extend to compute the transformed version of some empirical data. The sample data set with such a transformation applied is shown in Figure 2.3.

Given the transformed data we can easily calibrate the copula by maximising

$$l_c(\theta_c) = \sum_{t=1}^{T} \ln c(\bar{u}_1^{(t)}, \ldots, \bar{u}_d^{(t)}).$$

Sample Measure Approach Another approach, which is usually simpler to implement, is to calibrate to one or more sample dependence measures, such as $\bar{\tau}$ or $\bar{\rho}_s$. Of course this requires that we are able to compute the value of the dependence measure as a function of the copula parameters, and where possible we shall give expressions for τ and ρ_s for each of the models we describe in the next chapter. In order find a unique calibration we must have the number of equations equal to the number of unknown parameters.

We remark that the above approaches are not mutually exclusive, as we shall see in §3.1.2 where we calibrate using a mixture of the sample measure and maximum likelihood method. Note that some authors (e.g. [Che 04]) classify non-parametric copulas as a calibration technique, but we prefer to treat them as a model in their own right, which we discuss in detail in §4.

An important practical consideration is the presence of missing or otherwise bad data, which can frequently be a problem in calibrating to financial time series. Poor

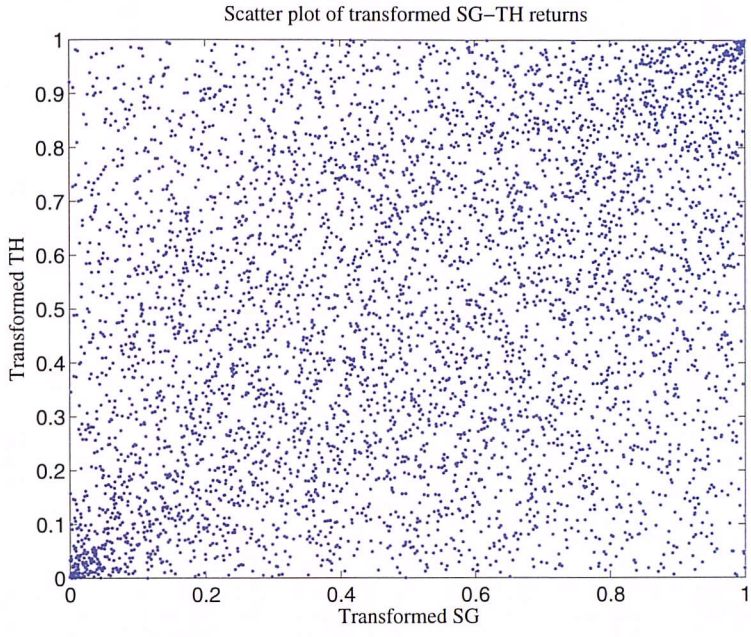


Figure 2.3: Empirical Marginal Transform of SG - TH returns

data quality makes the calibration of 'global' parameters – by which we mean those that impact all d dimensions of a copula – particularly difficult. To see this, suppose that we wish to calibrate to a d dimensional data set, and that a datum is 'bad' with probability p. Assuming that the event 'bad datum' occurs independently, the probability of any given day in the data set being completely free from bad data is $(1-p)^d$. The number of days with bad data grows linearly (to first order) in d. With $p = 2\%$ and $d = 20$, for example, we expect around a fifth of the data to be unusable. This problem is much less bad if we are able to calibrate parameters pairwise, since here we have that $d = 2$, regardless of the dimension of the data set.

2.3.2 Simulation

Monte Carlo simulation is widely used in quantitative finance when we wish to solve problems in multiple dimensions. Since it is in precisely these problems that copulas are useful, it is important to know how to generate random samples from the models we propose. The most general approach to this problem is based on the following result. The distribution of the k-th marginal u_k of some k dimensional copula C_k, conditional on $u_1, \ldots, u_{k-1}$ can be written as

$$P(U_k \leq u_k | U_1 = u_1 \ldots, U_{k-1} = u_{k-1}) = C_k(u_k | u_1, \ldots, u_{k-1})$$

$$= \frac{\partial^{k-1} C_k(u_1, \ldots, u_k)}{\partial u_1 \ldots \partial u_{k-1}} \Big/ \frac{\partial u_1 \ldots \partial u_{k-1}}{\partial^{k-1} C_{k-1}(u_1, \ldots, u_{k-1})},$$

where C_k is defined as $C_k = C(u_1, \ldots, u_k, 1, \ldots, 1)$ for $1 < k \leq d$. Thus we can construct an algorithm to generate a random sample $u_1, \ldots, u_d$ from an arbitrary copula C as follows.

General Simulation Algorithm

- Simulate u_1 from $U(0,1)$
- Simulate v from $U(0,1)$
- Set $u_2 = x$ where $C_2(x|u_1) = v$

$\vdots$

- Simulate v from $U(0,1)$
- Set $u_d = x$ where $C_d(x|u_1, \ldots, u_{d-1}) = v$

While this approach is very general, it can be very difficult to invert the function $C_k(x|u_1, \ldots, u_{k-1}) = v$ computationally, so it will usually be more efficient to implement an algorithm tailored to the copula we are using. We give examples of such algorithms for each of the copulas we consider in the next chapter.

Chapter 3

Parametric Copula Models

3.1 Existing Parametric Models

We now proceed to describe some of the main parametric copulas used in financial modelling. While our research efforts concentrate on developing new models of dependence, it is useful to examine the strengths and weaknesses of the existing parametric models as we will later compare them with our proposed models. There are essentially two approaches to constructing a parametric model: implicit copulas, which are those implied by an existing multivariate distribution; and explicit copulas, where the form of C is specified directly. We begin with some examples of the former.

3.1.1 Gaussian Copula

The Gaussian copula is used implicitly wherever we apply a multivariate normal distribution, and as such is undoubtedly the most widely used in quantitative finance.

Definition 11 *Given a d-dimensional standard multivariate normal distribution function Φ_d parameterised by a dispersion matrix $R \in [-1,1]^{d\times d}$, the* Gaussian copula *is the function $C_{\mathrm{Ga}} : [0,1]^d \to [0,1]$ such that*

$$C_{\mathrm{Ga}}(\mathbf{u}) = \Phi_d(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)); R).$$

Using (2.3) we have that the Gaussian copula has density

$$c_{\mathrm{Ga}}(\mathbf{u}) = \frac{\phi_d(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)); R)}{\prod_{i=1}^d \phi(\Phi^{-1}(u_i))} = \frac{1}{|R|^{1/2}} \exp\left[-\frac{1}{2}\zeta_\phi{}'(R^{-1} - I_{d\times d})\zeta_\phi\right]$$

where $I_{d\times d}$ denotes the $d \times d$ identity matrix and $\zeta_\phi = (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d))$. The bivariate density is plotted with

$$R = \begin{pmatrix} 1 & 0.32 \\ 0.32 & 1 \end{pmatrix}$$

in Figure 3.1 (NB: we have plotted this density and others on the grid $u_1 \times u_2 = (0.01, 0.03, \ldots, 0.99) \times (0.01, 0.03, \ldots, 0.99)$, since this and other copulas are not defined on the boundary $\partial[0,1]^2$).

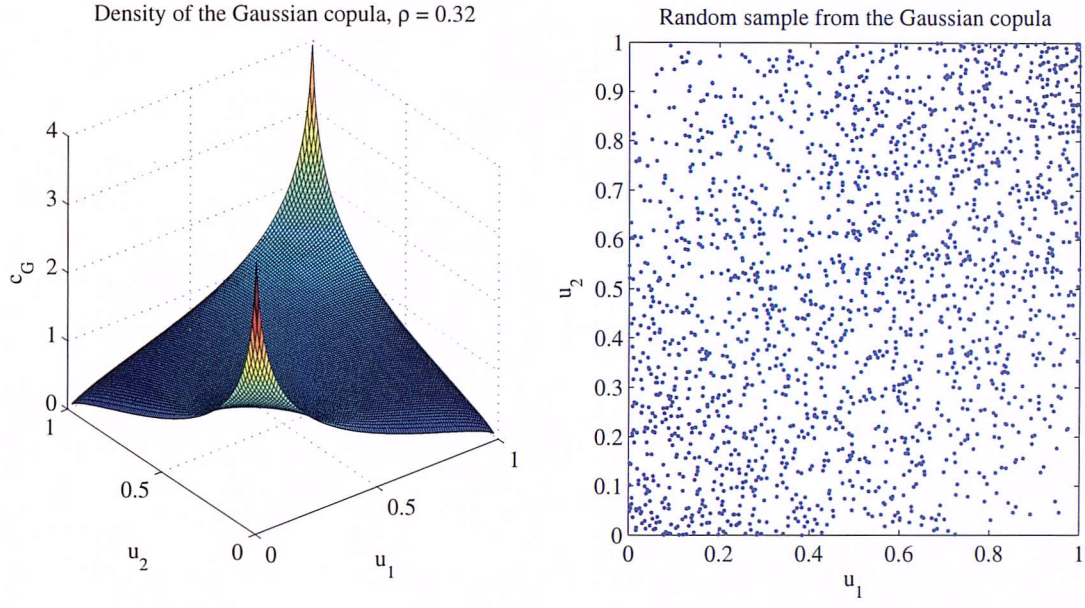


Figure 3.1: Gaussian copula

Perhaps the greatest strength of the Gaussian copula lies in its ease of application. It is completely parameterised by the dispersion matrix R, and where we have (or are willing to assume) Gaussian marginals we can calibrate directly to the sample correlation as in (2.6). Of course, the marginals will not in general be Gaussian, and in this case we cannot apply[1] (2.6) for the reasons described in §2.2.1. We could adopt a maximum likelihood approach, but a simpler solution is to make use of the relation

$$r = \sin\left(\frac{\pi\tau}{2}\right) \tag{3.1}$$

[1]It is important to understand the distinction between the matrix of linear correlations and the dispersion matrix, both of which are often referred to as 'the correlation matrix'. The value of the dispersion will be equal to the linear correlation if and only if the variables are elliptically distributed, i.e. they are joined by an elliptical copula *and* have elliptical marginals.

or

$$r = 2\sin\left(\frac{\pi\rho_s}{6}\right), \tag{3.2}$$

which hold for the Gaussian copula with any marginals. Since τ and ρ_s can easily be computed for an empirical sample, we have a simple method to calibrate the Gaussian copula. The calibration methodology is also pairwise, which is useful in practice as we mentioned in section §2.3.1 Using (3.1) with the sample data set we find that $r = 0.32$. Notice that this is different to the value of $\rho = 0.36$ computed by measuring correlation directly, an indication that the sample does not have a multivariate normal distribution.

Simulation from the Gaussian copula is also straightforward – we need only generate a sample from a standard normal distribution with correlation r, and transform it using the normal distribution function Φ.

Gaussian Simulation Algorithm [Emb 01]

- Compute the Cholesky decomposition of the dispersion matrix; $AA' = R$
- Simulate d independent $N(0,1)$ random variables, $z_1, \ldots, z_d$
- Set $x_i = A_{i1}z_1 + \cdots + A_{ii}z_i$ for $i = 1, \ldots, d$
- Set $u_i = \Phi(x_i)$ for $i = 1, \ldots, d$

2000 samples from the Gaussian copula with $r = 0.32$ are plotted in Figure 3.1.

Given the calibrated value of r, we can compute the tail dependence and quadrant dependence of the Gaussian copula. As is well known in the literature we have

$$\lambda_L = \lambda_U = 0$$

for $r \neq 1$, but we have previously argued that this of less interest than commonly thought. We can also compute ν_L and ν_U, which we plot in Figure 3.3.

Notice that the values for ν_L and ν_U are significantly smaller than those of the empirical data (see Figure 2.2), even though $\tau = \bar{\tau}$. The economic interpretation of this is that the probability of a large joint move is underestimated by the Gaussian copula model. Using a Gaussian copula can therefore to underestimation the risk in an equity portfolio, and so represents an important deficiency in this model. It is primarily this observation that has motivated the study of alternatives to the Gaussian dependence structure in quantitative finance.

3.1.2 Student's t copula and extensions

One of the most popular alternatives to the Gaussian model is the copula implied by multivariate Student's t distribution. The Student's t copula is defined in a similar way to the Gaussian:

Definition 12 *Given a d-dimensional multivariate t-distribution function $T_{\rho,\nu}$ parameterised by a dispersion matrix $R \in [-1,1]^{d\times d}$ and degree of freedom parameter $\nu \in \mathbf{R}$, the* Student's t copula *is the function $C_{\mathrm{t}} : [0,1]^d \to [0,1]$ such that*

$$C_{\mathrm{t}}(\mathbf{u}) = T_{\rho,\nu}(T_\nu^{-1}(u_1), \ldots, T_\nu^{-1}(u_d); R, \nu),$$

where $T_\nu^{-1}(\cdot)$ denotes the inverse of the univariate t distribution function having ν degrees of freedom.

The density function of the t copula can be expressed in a similar form to the Gaussian

$$\begin{aligned} c_{\mathrm{t}}(\mathbf{u}) &= \frac{t_d(T^{-1}(u_1), \ldots, T^{-1}(u_d); R, \nu)}{\prod_{i=1}^d t(T_\nu^{-1}(u_i); \nu)} \\ &= \frac{1}{|\rho|^{1/2}} \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left[\frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\right]^d \frac{\left(1+\frac{\zeta_t' \rho^{-1} \zeta_t}{\nu}\right)^{-\frac{\nu+d}{2}}}{\prod_{i=1}^d \left(1+\frac{\zeta_{ti}^2}{\nu}\right)^{-\frac{\nu+1}{2}}} \end{aligned}$$

where $\zeta_t = (T_\nu^{-1}(u_1), \ldots, T_\nu^{-1}(u_d))$.

The implementation of the t copula approach is not quite so simple as in the Gaussian case. We can easily estimate the dispersion matrix using (3.1) which has been shown to hold for all elliptical copulas by [Lin 03] (NB relation (3.2) applies only to the Gaussian copula) but the remaining parameter ν needs to be estimated using the maximum likelihood method. Such an approach was used by [Mas 02], who found values of ν in the range 7-12 for a variety of assets. Note that since the Gaussian copula can be shown to be equivalent to a t copula with $\nu \to \infty$, these results can be seen as a rejection of the Gaussian copula model for the assets studied. Applying this approach with $r = 0.32$ to the sample data we find that $\nu_{\mathrm{SG-TH}} = 7$. Notice that ν is a global parameter, which may create problems when this model with poor quality data sets.

Simulation from the t copula is not difficult – we need only generate samples from a multivariate t distribution with the appropriate parameters and then transform them to uniform (0,1) variates using the univariate Student's t distribution function T_ν (although numerical experiments show that this is somewhat slower than the Gaussian simulation algorithm, see Table 4.1 in §4.4.2).

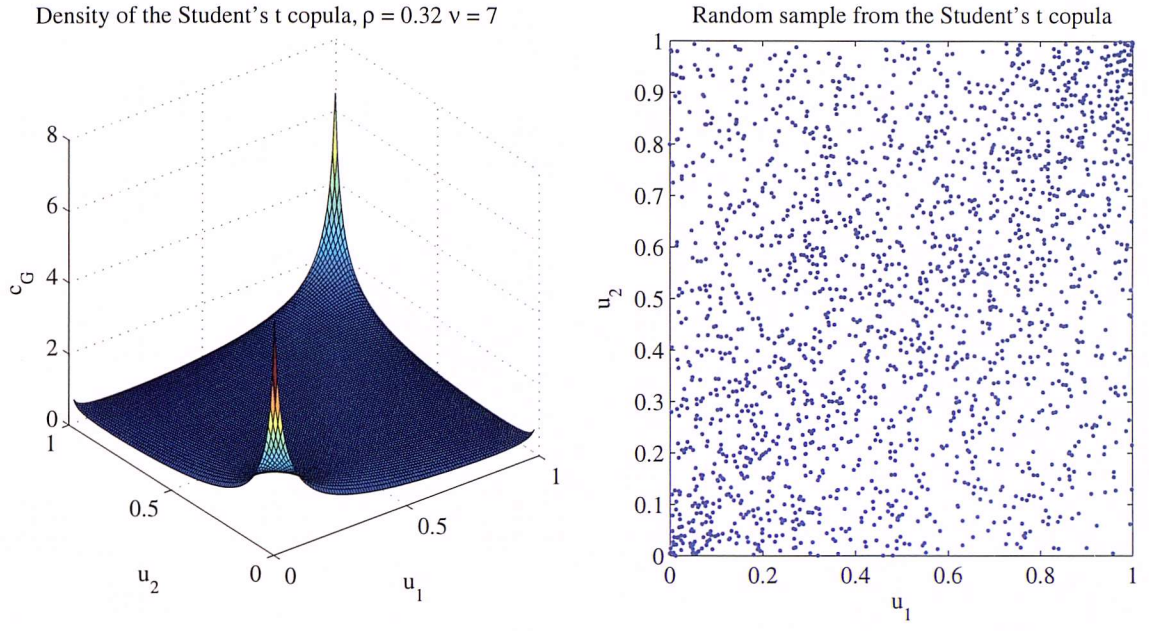


Figure 3.2: Student's t copula

Student's t Simulation Algorithm [Emb 01]

- Compute the Cholesky decomposition of the dispersion matrix; $AA' = R$
- Simulate d independent $N(0,1)$ random variables, $z_1, \ldots, z_d$
- Simulate a random variable s from the χ^2-distribution (independent of $z_1, \ldots, z_d$)
- Set $y_i = A_{i1}z_1 + \cdots + A_{ii}z_i$ for $i = 1, \ldots, d$
- Set $x_i = \sqrt{\frac{\nu}{s}}y_i$ for $i = 1, \ldots, d$
- Set $u_i = T_\nu(x_i)$ for $i = 1, \ldots, d$

2000 samples from the t copula with $r = 0.32$ and $\nu = 7$ are plotted in Figure 3.2.

The key difference between the t copula and the Gaussian is that the former exhibits greater dependence in the tails: using equation (2.7) we can show that, given our calibration,

$$\lambda_U = \lambda_L = 2T_{\nu+1}\left(-\sqrt{\nu+1}\frac{\sqrt{1-\rho}}{\sqrt{1+\rho}}\right).$$

The upper and lower quantile dependence functions are plotted in Figure 3.3.

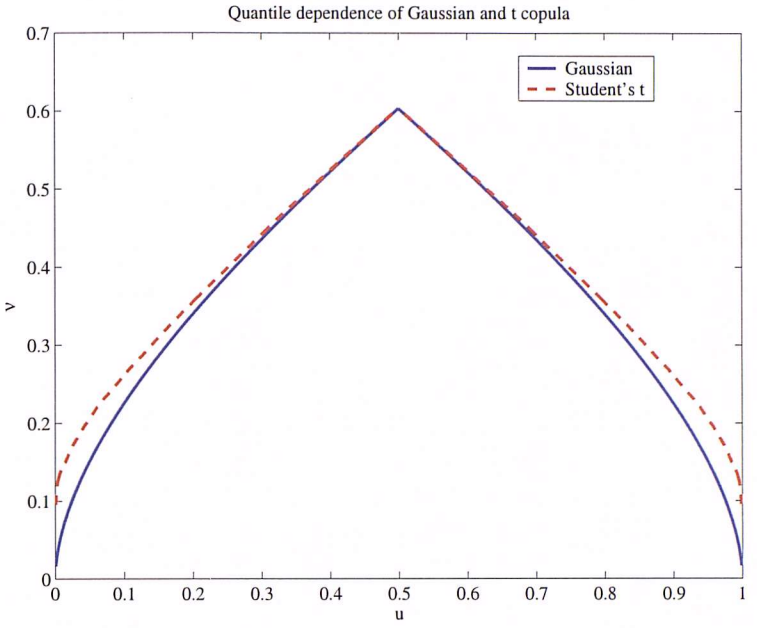


Figure 3.3: Quantile dependence of the Gaussian and Student's t copulas

We see that ν_L and ν_U are somewhat closer to their empirical values, so there is less likelihood of underestimating the risk. However, some limitations remain: the t copula, and indeed all elliptical copulas, exhibit strong symmetry that is not observed empirically. In particular they are pairwise radially symmetric, meaning that the upper and lower quadrant dependence will always be equal. As we have mentioned before, this is not consistent with empirical observation. Also, each pair of assets must have the same degree of freedom parameter, although they do of course have pairwise specific dispersion coefficient.

[Dem 05] propose a number of potential solutions by modifying the t copula to admit asymmetries. They suggest using a skewed t distribution to introduce asymmetries to the copula, but it is not clear how the parameters in this model could be calibrated in practice since the Kendall's tau relation (3.1) no longer holds. They also propose a grouped t copula, which has different degree of freedom parameter for each group of marginals. This model is easier to calibrate since (3.1) holds approximately (though the groups must be chosen *a priori*) but will be still radially symmetric. These authors also consider the tail copulas induced by the t distribution, but since we aim to find a model that can describe the whole dependency structure we consider them to be of less relevance to the problem in hand.

3.1.3 Archimedean Copulas

So far, we have considered copulas that are derived from existing multivariate distributions via Sklar's theorem. The class of Archimedean copulas arises differently – here we construct the copula directly using a function φ, known as a generator, which enables us to write the expression for the copula in closed form.

Definition 13 *Given a function* $\varphi : [0,1] \to [0,\infty)$ *such that that* $\varphi(1) = 0$ *and* $\varphi(0) = \infty$ *and having inverse* φ^{-1} *completely monotone, a d-dimensional* Archimedean copula *is a function* $C_\varphi : [0,1]^d \to [0,\infty)$ *such that*

$$C_\varphi(\mathbf{u}) = \varphi^{-1}(\varphi(u_1) + \ldots + \varphi(u_d)).$$

We also have that the general form for the density of a 2-dimensional Archimedean copula is

$$c_\varphi(u_1, u_2) = -\frac{\varphi''(C)\varphi'(u_1)\varphi'(u_2)}{\varphi'(C)^3}$$

The expression for a d-dimensional density could in principle be computed by repeated application the the chain rule (provided that derivatives of order d of φ exist), but no simple, general form is available.

The nature of an Archimedean copula is controlled by specifying a functional form for the generator which satisfies the properties in the definition. This means that one can construct a wide range of copulas in this family, see [Nel 99] §4 for a list of examples. While most of these are not appropriate for financial modelling applications, a couple of forms are frequently cited in the financial literature:

Definition 14 *Given a generator of the form* $\varphi(u) = u^{-\alpha} - 1$ *with* $\alpha \in (0,\infty)$ *the* Clayton copula *in d-dimensions is*

$$C_{\mathrm{Cl}}(u_1, \ldots, u_d) = (u_1^{-\alpha} + \cdots + u_d^{-\alpha} - d + 1)^{-\frac{1}{\alpha}}$$

Definition 15 *Given a generator of the form* $\varphi(u) = (-\ln(u))^\alpha$ *with* $\alpha \in (1,\infty)$ *the* Gumbel copula *in d-dimensions is*

$$C_{\mathrm{Gu}}(u_1, \ldots, u_d) = \exp\left\{-[(-\ln u_1)^\alpha + \cdots + (-\ln u_d)^\alpha]^{1/\alpha}\right\}$$

The attractiveness of Archimedean copulas from a modelling perspective comes from their asymmetry, see Figures 3.4 and 3.5 where we plot the density of each of the above together with 2000 random samples. In each case the copula parameter is chosen so that $\tau = 0.21$. This asymmetry is also reflected in their tail dependence coefficients, see

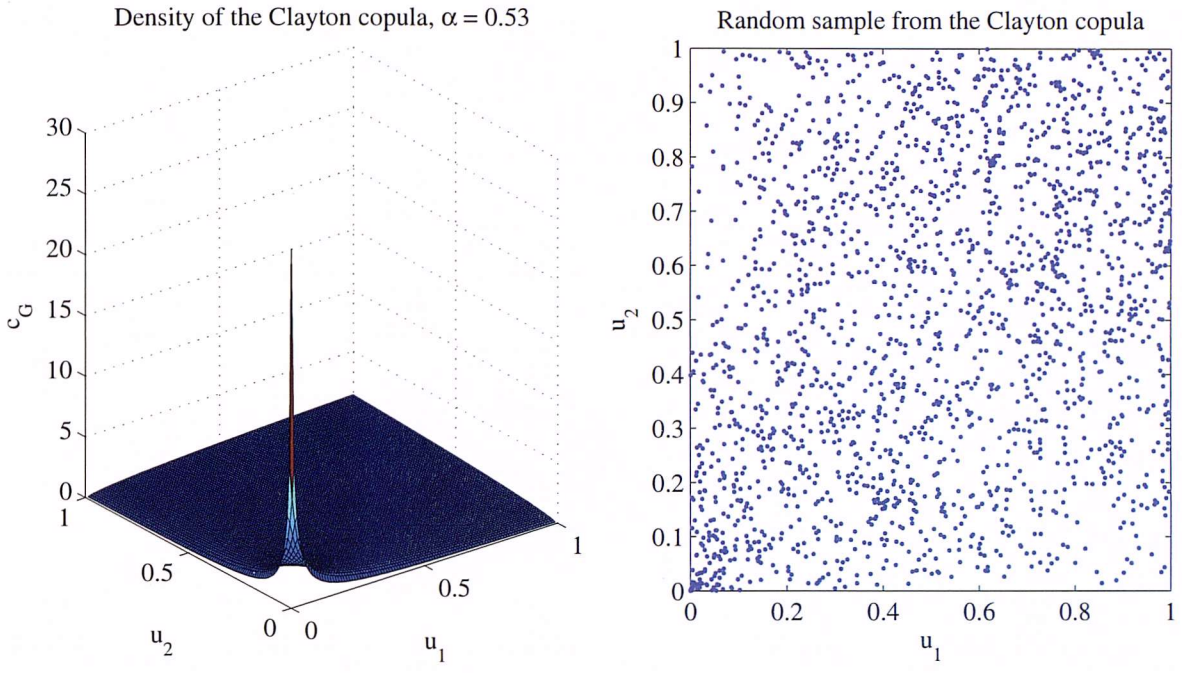


Figure 3.4: Clayton Copula

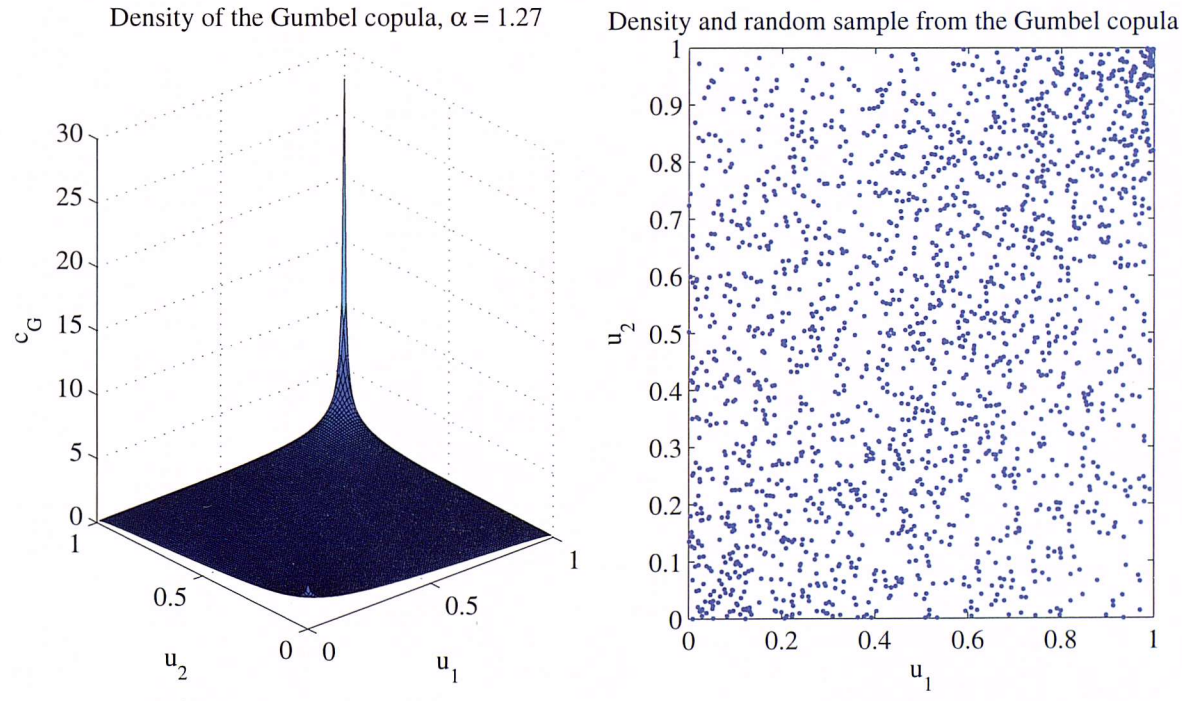


Figure 3.5: Gumbel Copula

Copula	τ	λ_U	λ_L
Clayton	$\frac{\alpha}{\alpha+2}$	0	$2^{-1/\alpha}$
Gumbel	$1-1/\alpha$	$2-2^{1/\alpha}$	0

Table 3.1: Parameters of the Clayton and Gumbel copulas

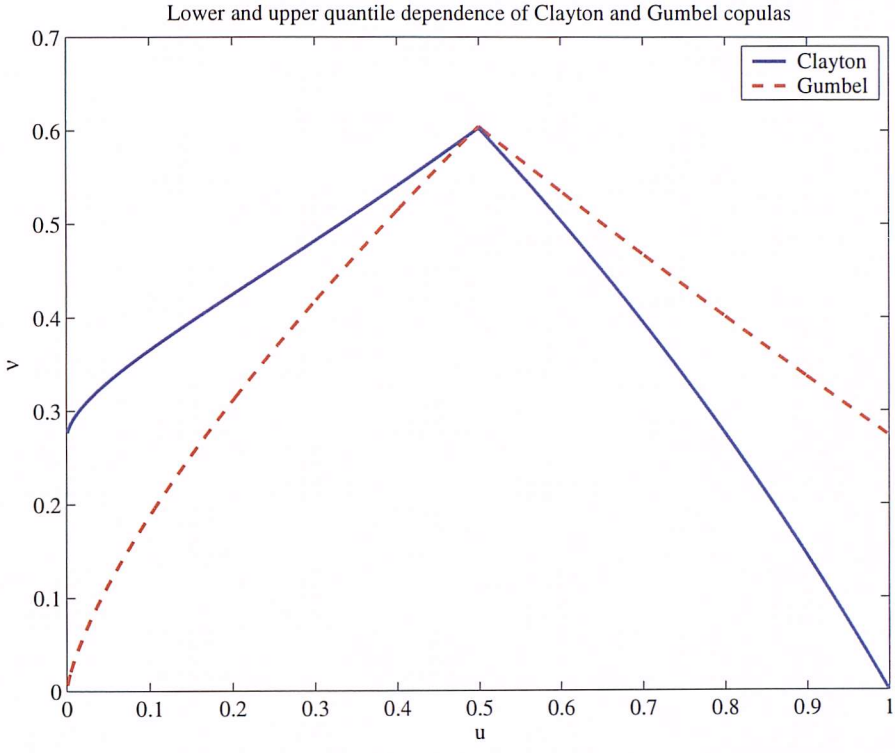


Figure 3.6: Quantile dependence of the Clayton and Gumbel copulas

Table 3.1, and in the lower and upper quantile dependence, which are plotted in Figure 3.6.

The Gumbel copula does not really exhibit the behaviour we want, in that $\nu_U > \nu_L$ while we have observed the reverse empirically. Let us then modify the model slightly and define the rotated Gumbel copula as follows:

Definition 16 *Given a generator of the form* $\varphi(u) = (-\ln(u))^\alpha$ *with* $\alpha \in (1, \infty)$ *the* rotated Gumbel copula *in d-dimensions is*

$$C_{\mathrm{Gu}}(u_1, \ldots, u_d) = u_1 + \cdots + u_d - 1 + \exp\left\{-\left[(-\ln u_1)^\alpha + \cdots + (-\ln u_d)^\alpha\right]^{1/\alpha}\right\}$$

For this copula we have

$$\tau = 1 - 1/\alpha \quad \lambda_L = 2 - 2^{1/\alpha} \quad \lambda_U = 0.$$

Efficient simulation techniques for Archimedean copulas are based on a result of [Mar 88] that, for certain Archimedean copulas, φ^{-1} can be written as the inverse Laplace transform of some distribution function $F_{\varphi^{-1}}[0,\infty) \to [0,\infty)$, where $F_{\varphi^{-1}}(0) = 0$. With this interpretation, a general simulation algorithm for Archimedean copulas can be constructed.

Archimedean Copula Simulation Algorithm

- Simulate d independent uniform random variables, $v_1, \ldots, v_d$
- Simulate a random variable Y from the distribution of $F_{\varphi^{-1}}$
- Set $\vartheta_i = \frac{-\ln(v_i)}{Y}$
- Set $u_i = \varphi^{-1}(\vartheta_i)$

See [Mel 06] for details of how to implement this algorithm for the Clayton and Gumbel copulas.

Bivariate Archimedean copulas are easy to calibrate, since the parameter α can be expressed as a function of τ (see table). For the sample data we find

$$\alpha_{\mathrm{Cl}} = 0.53 \quad \text{and} \quad \alpha_{\mathrm{Gu}} = 1.27$$

It is more difficult to calibrate Archimedean copulas in more than two dimensions. One quick and easy approach, which we use in our implementation in §5.1, is to choose α based on the median of the off-diagonal entries in the matrix of τ.

A more sophisticated approach is described by [Gen 93], but we are still constrained by having just one free parameter α. Indeed, the applicability of these and other Archimedean copulas would appear to be severely limited by their lack of parameters – each copula is parameterised by a single number α, regardless of its dimensionality. It is difficult to see how the d-dimensional dependency structure of a possibly diverse range of assets can successfully be characterised by one parameter for even moderate d, which would seem to preclude the use of such copulas in practice in most quantitative finance applications. Also, while empirical data are often observed to be asymmetric, the degree of asymmetry is not so great as that exhibited by the Clayton and Archimedean copulas (compare Figure 2.2 and Figure 3.6).

3.1.4 Mixture Copulas

Some authors have recently tried to resolve the problems described above by using mixtures of parametric copulas (e.g. [Hu 03]). This approach is based on the observation that, given some functions $C_1, \ldots, C_M$, each satisfying the definition of a copula in d-dimensions, and some weight parameters $w_1, \ldots, w_M$ such that $\sum_{i=1}^{M} w_i = 1$, the function

$$C(\mathbf{u}) = w_1 C_1(\mathbf{u}) + \cdots + w_M C_M(\mathbf{u}) \tag{3.3}$$

also satisfies the definition of a copula. This enables one to construct a new copula which captures a wide range of dependence behaviour – [Hu 03], for example, uses combinations of Gaussian, Gumbel and rotated Gumbel copulas to model dependence between stock market indexes. Note that this approach broadly belongs to the class of implicit approaches, since we work with mixtures of copulas directly. In the next section we shall construct an implicit mixture model based on the multivariate mixture of normals.

Mixture copulas are very flexible and simulation is as easy as it is for each of their components, but the main practical problem will be parameter estimation – we need to estimate $M-1$ weight parameters plus all of the parameters for the component copulas, the number of which will grow quickly as d increases. Moreover, it is not clear how to estimate the dependence measures τ and ρ_s for mixture copulas, meaning that the calibration must be done by numerically cumbersome maximum likelihood techniques.

3.2 Gaussian Mixture Copula

3.2.1 Definitions

We introduce the notion of the Gaussian mixture copula, that is to say the copula induced by a Gaussian mixture distribution.

Definition 17 *Given M multivariate normal distribution functions each parameterised*[2] *by mean vectors $\mu^{(k)} \in \mathbf{R}^d$, positive definite correlation matrices $R^{(k)} \in [-1,1]^{d\times d}$, variance vectors $\sigma^{(k)} \in [0,\infty)^d$ and probabilities $p^{(k)}$ such that $\sum_{k=1}^{M} p^{(k)} = 1$, we define the Gaussian Mixture distribution via its cdf as*

$$F_{\mathrm{GM}}(\mathbf{x}) \equiv \sum_{k=1}^{M} p^{(k)} \Phi(\mathbf{x}; \mu^{(k)}, R^{(k)}, \sigma^{(k)}).$$

It will sometimes be more concise to express the GM distribution in terms of the dispersion matrices

$$\Sigma_{ij}^{(k)} = \sigma_i^{(k)} R_{ij}^{(k)} \sigma_j^{(k)} \qquad \text{for} \quad i,j = 1,\ldots,d.$$

Clearly, the pdf of this distribution is

$$f_{\mathrm{GM}}(\mathbf{x}) = \sum_{k=1}^{M} p^{(k)} \phi(\mathbf{x}; \mu^{(k)}, R^{(k)}, \sigma^{(k)}).$$

It can be shown ([McL 00]) that the GM distribution has mean

$$\mu_i = E[X_{\mathrm{GM}_i}] = \sum_{k=1}^{M} p^{(k)} \mu_i^{(k)}$$

and covariance

$$\sigma_{ij} = E[X_{\mathrm{GM}_i} X_{\mathrm{GM}_j}] = \sum_{k=1}^{M} p^{(k)} \Sigma_{ij}^{(k)} + \sum_{k=1}^{M} p^{(k)} (\mu_i^{(k)} - \mu_i)(\mu_j^{(k)} - \mu_j).$$

In what follows we shall assume without loss of generality that the GM distribution has zero mean and unit standard deviation in each dimension. Expressions for the Kendall's tau and Spearman's rho of this distribution are derived in the following section.

[Fan 87] show that this distribution is elliptical if and only if $\mu^{(k)} = \mu$, $R^{(k)} = R$ and for all $k = 1,\ldots,M$ and $\sigma^{(k)} = n\sigma^{(l)}$ with $n \in \mathbf{R}$ for all $k,l = 1,\ldots,M$. The radial symmetry of this distribution has been investigated by [Dat 02], who makes the following definition.

[2] In this section the notation unavoidably becomes rather heavily loaded. It may assist the reader to note that we shall use superscript indices $^{(k)}$,$^{(l)}$ and $^{(m)}$ to denote one of the M Gaussians in the mixture, and use subscript indices $_i$ and $_j$ to denote one of the d dimensions.

Definition 18 *Let S denote the set of all distinct dispersion matrices $\Sigma^{(k)}$, $k = 1, \ldots, M$ given by $S = \{\Delta^{(1)}, \ldots, \Delta^{(N)}\}$ $(N \leq M)$. We denote the indices of these matrices as A_α, the means as M_α and the weights as W_α, i.e.*

$$A_\alpha = \{k : \Sigma^{(k)} = \Delta^{(\alpha)}\}, \quad M_\alpha = \{\mu^{(k)} : k \in A_\alpha\}, \quad W_\alpha = \sum_{j \in A_\alpha} w_j$$

where $\alpha = 1, \ldots, N$. Let $Q_1, \ldots, Q_N$ each be random vectors having the 'distribution of the means of S', i.e. each Q_α is a discrete random vector that has probability distribution

$$P\left(Q_\alpha = \mu^{(k)}\right) = \frac{w_k}{W_\alpha} \quad \text{for } k \in A_\alpha.$$

Theorem 6 *A Gaussian Mixture distribution is symmetric about $[a, b]$ if and only if the distribution of each $Q_1, \ldots, Q_N$ is symmetric about $[a, b]$.*

Proof See [Dat 02].

We can now apply the corollary to Sklar's theorem to implicity define the Gaussian mixture copula

Definition 19 *Given a d-dimensional Gaussian mixture distribution F_{GM}, the* Gaussian Mixture copula *is the function $C_{\text{GM}} : [0, 1]^d \to [0, 1]$ such that*

$$C_{\text{GM}}(\mathbf{u}) = F_{\text{GM}}(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))$$

where $F_{\text{GM}_i}(x)$ denotes the i-th marginal distribution, i.e.

$$F_{\text{GM}_i}(x) = \sum_{k=1}^{M} p^{(k)} \Phi(x_i; \mu_i^{(k)}, \sigma_i^{(k)}).$$

We can also define the Gaussian mixture copula density as

$$c_{\text{GM}}(\mathbf{u}) = \frac{f_{\text{GM}}(\mathbf{x})}{\prod_{i=1}^{d} f_{\text{GM}_i}(x_i)}$$

where $x_i = F_{\text{GM}_i}^{-1}(u_i)$ and

$$f_{\text{GM}_i}(x) = \sum_{k=1}^{M} p^{(k)} \phi(x; \mu_i^{(k)}, \Sigma_i^{(k)}).$$

Thus we can always write the GM copula in terms of standard Gaussian distributions and their inverses. We hope it is clear that the Gaussian Mixture copula defined in this way is not equivalent to the class of mixture copula defined by (3.3): here we

have implicitly defined a copula via a mixture of multivariate distributions, rather than simply summing existing copula functions, and we aim to show that our approach leads a flexible yet tractable model of dependency. The three component bivariate GM copula density with parameter set

$$p^{(1)} = 0.6 \qquad p^{(2)} = 0.2 \qquad p^{(3)} = 1 - p^{(1)} - p^{(2)} = 0.2$$

$$\mu_1^{(1)} = \mu_2^{(1)} = 0 \qquad \mu_1^{(2)} = \mu_2^{(2)} = -1 \qquad \mu_1^{(3)} = \mu_2^{(3)} = 1$$

$$\sigma_1^{(1)} = \sigma_2^{(1)} = 1 \qquad \sigma_1^{(2)} = \sigma_2^{(2)} = 1 \qquad \sigma_1^{(3)} = \sigma_2^{(3)} = 1$$

$$r^{(1)} = -0.21 \qquad r^{(2)} = 0.67 \qquad r^{(3)} = 0.44$$

is plotted in Figure 3.7. Clearly the GM copula does not suffer from a lack of parameters – on the contrary there are rather too many parameters to be calibrated easily. We consider the question of reducing the number of parameters while retaining the ability to model empirical data well in section §3.2.3.

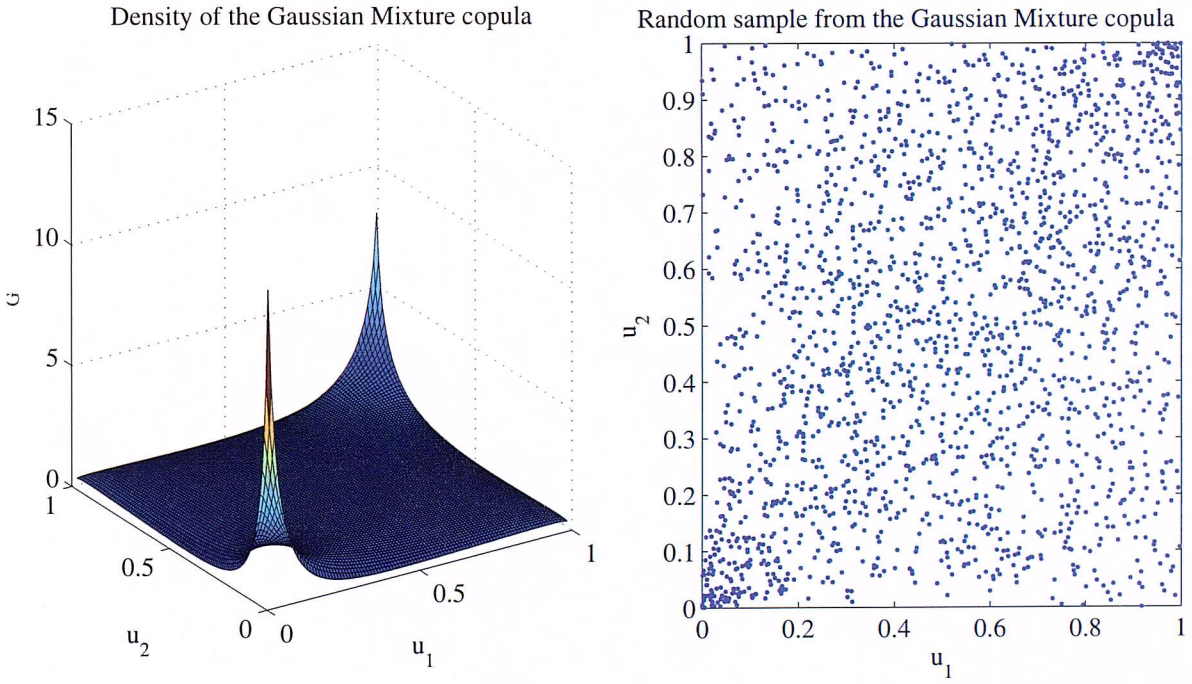


Figure 3.7: Gaussian Mixture copula

Notice from Figure 3.7 that the GM copula is not radially symmetric.

Theorem 7 *A GM copula is radially symmetric if and only if the distribution of Q_α is symmetric about $[a, b]$ for every $\alpha = 1, \ldots, N$.*

Proof This follows immediately from Theorem 5 and Theorem 6.

Clearly the GM copula does not suffer from the excessive symmetry exhibited by the Gaussian and t copula models, and is much more parameter rich than the Archimedean copulas.

3.2.2 Dependence Measures

The next task is to derive analytic expressions for the dependence measures of this copula. First let us consider Kendall's tau.

Theorem 8 Kendall's Tau for GM copula *Let C_{GM} be a d-dimensional Gaussian Mixture copula with parameters $\mu^{(k)}$, $R^{(k)}$, $\sigma^{(k)}$ and $p^{(k)}$ for $k = 1, \ldots, M$. Then the Kendall's tau matrix T is*

$$T_{ij} = 2 \times \sum_{k=1}^{M} \sum_{l=1}^{M} p^{(k)} p^{(l)} (\Phi_{ij}^{(k,l)^-} + \Phi_{ij}^{(k,l)^+}) - 1 \qquad \text{for} \quad i, j = 1, \ldots, d. \tag{3.4}$$

Here $\Phi_{ij}^{(k,l)^-}$ and $\Phi_{ij}^{(k,l)^+}$ are functions of the copula parameters:

$$\Phi_{ij}^{(k,l)^-} = \Phi\left(\frac{-\mu_i^{(k,l)}}{\sigma_i^{(k,l)}}, \frac{-\mu_j^{(k,l)}}{\sigma_j^{(k,l)}}; R_{ij}^{(k,l)}\right) \qquad \Phi_{ij}^{(k,l)^+} = \Phi\left(\frac{\mu_i^{(k,l)}}{\sigma_i^{(k,l)}}, \frac{\mu_j^{(k,l)}}{\sigma_j^{(k,l)}}; R_{ij}^{(k,l)}\right)$$

where

$$\mu_i^{(k,l)} = \mu_i^{(k)} - \mu_i^{(l)}, \qquad \mu_j^{(k,l)} = \mu_j^{(k)} - \mu_j^{(l)}, \qquad \sigma_i^{(k,l)} = \sqrt{\sigma_i^{(k)^2} + \sigma_i^{(l)^2}},$$

$$\sigma_j^{(k,l)} = \sqrt{\sigma_j^{(k)^2} + \sigma_j^{(l)^2}} \qquad \text{and} \qquad R_{ij}^{(k,l)} = \frac{R_{ij}^{(k)} \sigma_i^{(k)} \sigma_j^{(k)} + R_{ij}^{(l)} \sigma_i^{(l)} \sigma_j^{(l)}}{\sqrt{\sigma_i^{(k)^2} + \sigma_i^{(l)^2}} \sqrt{\sigma_j^{(k)^2} + \sigma_j^{(l)^2}}}.$$

Proof Given two bivariate samples (X_i, X_j) and $(\acute{X}_i, \acute{X}_j)$ drawn independently from the ith and jth dimensions of the copula C, Kendall's tau is defined as

$$T_{ij} = P[(X_i - \acute{X}_i)(X_j - \acute{X}_j) > 0] - P[(X_i - \acute{X}_i)(X_j - \acute{X}_j) < 0].$$

Now consider a pair of samples from our copula, $\mathbf{X} = (X_i, X_j)$ and $\acute{\mathbf{X}} = (\acute{X}_i, \acute{X}_j)$, and let

$$Z_i = X_i - \acute{X}_i \qquad \text{and} \qquad Z_j = X_j - \acute{X}_j$$

It is clear that

$$T_{ij} = P[Z_i Z_j > 0] - P[Z_i Z_j < 0] = 2P[Z_i Z_j > 0] - 1$$

$$= 2P[Z_i > 0 \cap Z_j > 0] + 2P[Z_i < 0 \cap Z_j < 0] - 1. \tag{3.5}$$

Hence computing T_{ij} is reduced to estimating the joint distribution of (Z_i, Z_j). We have that (X_i, X_j) is sampled from the kth of the M normals with probability $p^{(k)}$, and $(\acute{X}_i, \acute{X}_j)$ is sampled independently from the lth distribution with probability $p^{(l)}$. Let us write $(X_i^{(k)}, X_j^{(k)})$ to denote that $\mathbf{X}$ is sampled from the kth normal, and similarly for $\acute{\mathbf{X}}$. We can then write that

$$Z_i^{(k,l)} = X_i^{(k)} - \acute{X}_i^{(l)} \qquad Z_j^{(k,l)} = X_j^{(k)} - \acute{X}_j^{(l)}$$

The key observation is that, as sums of Gaussian random variables, $(Z_i^{(k,l)}, Z_j^{(k,l)})$ are themselves bivariate Gaussian distributed. Applying some standard results on the properties of the sum of independent random variables we have that these have means

$$\mu_i^{(k,l)} \equiv E[Z_i^{(k,l)}] = \mu_i^{(k)} - \mu_i^{(l)} \qquad \mu_j^{(k,l)} \equiv E[Z_j^{(k,l)}] = \mu_j^{(k)} - \mu_j^{(l)},$$

variances

$$\sigma_i^{(k,l)^2} \equiv E[Z_i^{(k,l)^2}] = \sigma_i^{(k)^2} + \sigma_i^{(l)^2} \qquad \sigma_j^{(k,l)^2} \equiv E[Z_j^{(k,l)^2}] = \sigma_j^{(k)^2} + \sigma_j^{(l)^2}$$

and correlation

$$r_{ij}^{(k,l)} \equiv \frac{\mathrm{Cov}[Z_i^{(k,l)}, Z_j^{(k,l)}]}{\sqrt{\mathrm{Var}[Z_i^{(k,l)}]\mathrm{Var}[Z_j^{(k,l)}]}} = \frac{r_{ij}^{(k)}\sigma_i^{(k)}\sigma_j^{(k)} + r_{ij}^{(l)}\sigma_i^{(l)}\sigma_j^{(l)}}{\sqrt{\sigma_i^{(k)^2} + \sigma_i^{(l)^2}}\sqrt{\sigma_j^{(k)^2} + \sigma_j^{(l)^2}}}.$$

So, given that we sample from the kth and lth distributions, the required probabilities are

$$\Phi_{ij}^{(k,l)^-} \equiv P[Z_i^{(k,l)} < 0 \cap Z_j^{(k,l)} < 0] = \Phi_2\left(\frac{-\mu_i^{(k,l)}}{\sigma_i^{(k,l)}}, \frac{-\mu_j^{(k,l)}}{\sigma_j^{(k,l)}}; r_{ij}^{(k,l)}\right)$$

$$\Phi_{ij}^{(k,l)^+} \equiv P[Z_i^{(k,l)} > 0 \cap Z_j^{(k,l)} > 0] = \Phi_2\left(\frac{\mu_i^{(k,l)}}{\sigma_i^{(k,l)}}, \frac{\mu_j^{(k,l)}}{\sigma_j^{(k,l)}}; r_{ij}^{(k,l)}\right),$$

where $\Phi_2(x, y; r)$ denotes the cdf of the standard bivariate Gaussian distribution with correlation r evaluated at the point (x, y).

Now

$$P[Z_i < 0 \cap Z_j < 0] = \sum_k \sum_l P[Z_i < 0 \cap Z_j < 0 | \text{sample from } k, l] P[\text{sample from } k, l]$$

$$= \sum_k \sum_l P[Z_i^{(k,l)} < 0 \cap Z_j^{(k,l)} < 0] p^{(k)} p^{(l)} = \sum_k \sum_l p^{(k)} p^{(l)} \Phi_{ij}^{(k,l)^-}$$

and similarly

$$P[Z_i > 0 \cap Z_j > 0] = \sum_k \sum_l p^{(k)} p^{(l)} \Phi_{ij}^{(k,l)^+}.$$

Substituting into (3.5) we have the required result

$$T_{ij} = 2 \times \sum_{k=1}^{M} \sum_{l=1}^{M} p^{(k)} p^{(l)} (\Phi_{ij}^{(k,l)^-} + \Phi_{ij}^{(k,l)^+}) - 1.$$

$\diamondsuit$

Remark Given the special case that the Gaussian Mixture distribution is elliptical we may use the result of [Lin 03] that

$$T_{ij} = \frac{2}{\pi} \arcsin(r_{ij}).$$

While the expression in 3.4 may appear rather complex, it is simply the sum of bivariate normal distribution functions[3] applied to the (known) parameters of the copula, and can thus can be computed quickly and easily.

A similar approach can be used to compute Spearman's rho for the GM copula

Theorem 9 Spearman's Rho for GM copula *Let C_{GM} be a d-dimensional Gaussian Mixture copula with parameters $\mu^{(k)}$, $R^{(k)}$, $\sigma^{(k)}$ and $p^{(k)}$ for $k = 1, \ldots, M$. Then the Spearman's rho matrix is*

$$R_{Sij} = 6 \times \sum_{k=1}^{M} \sum_{l=1}^{M} \sum_{m=1}^{M} p^{(k)} p^{(l)} p^{(m)} (\Phi_{ij}^{(k,l,m)^-} + \Phi_{ij}^{(k,l,m)^+}) - 3. \tag{3.6}$$

Here $\Phi_{ij}^{(k,l,m)^-}$ and $\Phi_{ij}^{(k,l,m)^+}$ are functions of the copula parameters:

$$\Phi_{ij}^{(k,l,m)^-} = \Phi\left(\frac{-\mu_i^{(k,l)}}{\sigma_i^{(k,l)}}, \frac{-\mu_j^{(k,m)}}{\sigma_j^{(k,m)}}; r_{ij}^{(k,l,m)}\right) \qquad \Phi_{ij}^{(k,l,m)^+} = \Phi\left(\frac{\mu_i^{(k,l)}}{\sigma_i^{(k,l)}}, \frac{\mu_j^{(k,m)}}{\sigma_j^{(k,m)}}; r_{ij}^{(k,l,m)}\right)$$

where

$$\mu_i^{(k,l)} = \mu_i^{(k)} - \mu_i^{(l)} \qquad \mu_j^{(k,m)} = \mu_j^{(k)} - \mu_j^{(m)}, \qquad \sigma_i^{(k,l)} = \sqrt{\sigma_i^{(k)^2} + \sigma_i^{(l)^2}}$$

$$\sigma_j^{(k,m)} = \sqrt{\sigma_j^{(k)^2} + \sigma_j^{(m)^2}} \qquad \text{and} \qquad r_{ij}^{(k,l,m)} = \frac{r_{ij}^{(k)} \sigma_i^{(k)} \sigma_j^{(k)}}{\sqrt{\sigma_i^{(k)^2} + \sigma_i^{(l)^2}} \sqrt{\sigma_j^{(k)^2} + \sigma_j^{(m)^2}}}.$$

[3]Computing the value of this function numerically, which will be central to our implementation methodology, is not entirely trivial but see [Wes 05] for details of a fast double precision algorithm.

Proof We can apply a method similar to that used for Kendall's tau to compute the Spearman's rho of the GM copula. Consider pairs of samples from our copula, $\mathbf{X} = (X_i, X_j)$, $\acute{\mathbf{X}} = (\acute{X}_i, \acute{X}_j)$ and $\grave{\mathbf{X}} = (\grave{X}_i, \grave{X}_j)$, and let

$$Z_i = X_i - \acute{X}_i \qquad \text{and} \qquad Z_j = X_j - \grave{X}_j$$

Now, it is clear that

$$R_{Sij} = 3 \times (P[Z_i Z_j > 0] - P[Z_i Z_j < 0]) = 6P[Z_i Z_j > 0] - 3$$

$$= 6P[Z_i > 0 \cap Z_j > 0] + 6P[Z_i < 0 \cap Z_j < 0] - 3. \tag{3.7}$$

Hence computing ρ_S is again reduced to estimating the joint distribution of (Z_i, Z_j). We have that (X_i, X_j) is sampled from the kth of the M normals with probability $p^{(k)}$, that $(\acute{X}_i, \acute{X}_j)$ is sampled from the lth distribution with probability $p^{(l)}$ and that $(\grave{X}_i, \grave{X}_j)$ is sampled from the mth distribution with probability $p^{(m)}$. Let us then write $(X_i^{(k)}, X_j^{(k)})$ to denote that $\mathbf{X}$ is sampled from the kth normal, and similarly for $\acute{\mathbf{X}}$ and $\grave{\mathbf{X}}$. We can then write that

$$Z_i^{(k,l)} = X_i^{(k)} - \acute{X}_i^{(l)} \qquad Z_j^{(k,m)} = X_j^{(k)} - \grave{X}_j^{(m)}$$

Again, the key observation is that $(Z_i^{(k,l)}, Z_j^{(k,m)})$ are themselves bivariate Gaussian distributed. Applying some standard results on the properties of the sum of independent random variables we see that these have means

$$\mu_i^{(k,l)} \equiv E[Z_i^{(k,l)}] = \mu_i^{(k)} - \mu_i^{(l)} \qquad \mu_j^{(k,m)} \equiv E[Z_j^{(k,m)}] = \mu_j^{(k)} - \mu_j^{(m)},$$

variances

$$\sigma_i^{(k,l)^2} \equiv E[Z_i^{(k,l)^2}] = \sigma_i^{(k)^2} + \sigma_i^{(l)^2} \qquad \sigma_j^{(k,m)^2} \equiv E[Z_j^{(k,m)^2}] = \sigma_j^{(k)^2} + \sigma_j^{(m)^2}$$

and correlation

$$r_{ij}^{(k,l,m)} \equiv \frac{\mathrm{Cov}[Z_i^{(k,l)}, Z_j^{(k,m)}]}{\sqrt{\mathrm{Var}[Z_i^{(k,l)}]\mathrm{Var}[Z_j^{(k,m)}]}} = \frac{r_{ij}^{(k)}\sigma_i^{(k)}\sigma_j^{(k)}}{\sqrt{\sigma_i^{(k)^2} + \sigma_i^{(l)^2}}\sqrt{\sigma_j^{(k)^2} + \sigma_j^{(m)^2}}}.$$

The latter expression is different from that in the expression for Kendall's tau since in this case $\acute{X}_i^{(l)}$ and $\grave{X}_j^{(m)}$ are independent, so $r_{ij}^{(l,m)} = 0$. Now, given that we sample from the kth, lth and mth distributions, the required probabilities are

$$\Phi_{ij}^{(k,l,m)^-} \equiv P[Z_i^{(k,l)} < 0 \cap Z_j^{(k,m)} < 0] = \Phi_2\left(\frac{-\mu_i^{(k,l)}}{\sigma_i^{(k,l)}}, \frac{-\mu_j^{(k,m)}}{\sigma_j^{(k,m)}}; r_{ij}^{(k,l,m)}\right)$$

$$\Phi_{ij}^{(k,l,m)^+} \equiv P[Z_i^{(k,l)} > 0 \cap Z_j^{(k,m)} > 0] = \Phi_2\left(\frac{\mu_i^{(k,l)}}{\sigma_i^{(k,l)}}, \frac{\mu_j^{(k,m)}}{\sigma_j^{(k,m)}}; r_{ij}^{(k,l,m)}\right).$$

Now

$$P[Z_i < 0 \cap Z_j < 0] = \sum_k \sum_l \sum_m P[Z_i < 0 \cap Z_j < 0 | \text{sample from } k,l,m] P[\text{sample from } k,l,m]$$

$$= \sum_k \sum_l \sum_m P[Z_i^{(k,l)} < 0 \cap Z_j^{(k,m)} < 0] p^{(k)} p^{(l)} p^{(m)} = \sum_k \sum_l \sum_m p^{(k)} p^{(l)} p^{(m)} \Phi_{ij}^{(k,l,m)^-}$$

and similarly

$$P[Z_i > 0 \cap Z_j > 0] = \sum_k \sum_l \sum_m p^{(k)} p^{(l)} p^{(m)} \Phi_{ij}^{(k,l,m)^+}.$$

Substituting into (3.7) we have the required result

$$R_{Sij} = 6 \times \sum_{k=1}^{M} \sum_{l=1}^{M} \sum_{l=1}^{M} p^{(k)} p^{(l)} p^{(m)} (\Phi_{ij}^{(k,l,m)^-} + \Phi_{ij}^{(k,l,m)^+}) - 3.$$

$\Diamond$

Note that while we have

$$\rho_{Sij} = \frac{6}{\pi} \arcsin(\frac{r_{ij}}{2})$$

for the standard Gaussian distribution, this does *not* hold for general elliptical distributions, including the special case where the GM distribution is elliptical.

We remark that while we have focussed on the dependence properties of the copula, the results also apply to the GM distribution itself. To the best of our knowledge the above theorems are the first available expressions for the non-parametric dependence of this widely used statistical model.

Let us now examine the tail dependence of the GM copula. We require some auxiliary definitions before we can present our main result.

Definition 20 *Suppose we have a bivariate GM copula with parameters denoted as above. If there exist components of the mixture that have greatest variance in both dimensions, then we say they are* tail dominant *and we denote them by* k^**, i.e.*

$$\{k^* = k : \sigma_1^{(k^*)} \geq \sigma_1^{(k)} \text{ and } \sigma_2^{(k^*)} \geq \sigma_2^{(k)} \text{ for all } k = 1, \ldots, M\}.$$

If there components that are tail dominant and have mean lower than all other tail dominant components in both dimensions, then we say they are lower tail dominant *and denote them* k^-*, i.e.*

$$\{k^- = k^* : \mu_1^{(k^-)} \leq \mu_1^{(k^*)} \text{ and } \mu_2^{(k^-)} \leq \mu_2^{(k^*)} \text{ for all } k^*\}.$$

Similarly, if there components that are tail dominant and have mean lower than all other tail dominant components in both dimensions we say they are upper tail dominant *and denote them* k^+*, i.e.*

$$\{k^+ = k^* : \mu_1^{(k^+)} \geq \mu_1^{(k^*)} \text{ and } \mu_2^{(k^+)} \geq \mu_2^{(k^*)} \text{ for all } k^*\}.$$

Clearly, if a GM copula has just one tail dominant component that component will be both lower and upper tail dominant.

Theorem 10 *A bivariate copula with lower tail dominant components* k^- *has*

$$\lambda_L = \frac{\sum_{k \in k^-} p^{(k)} \mathbf{1}_{[r^{(k)}=1]}}{\sum_{k \in k^-} p^{(k)}}.$$

If a GM copula has no lower tail dominant distributions, i.e. if $k^- = \emptyset$*, then*

$$\lambda_L = 0.$$

A bivariate copula with upper tail dominant components k^+ *has*

$$\lambda_U = \frac{\sum_{k \in k^+} p^{(k)} \mathbf{1}_{[r^{(k)}=1]}}{\sum_{k \in k^+} p^{(k)}}.$$

If a GM copula has no upper tail dominant distributions, i.e. if $k^+ = \emptyset$*, then*

$$\lambda_U = 0.$$

Proof We prove the result for λ_L. It has been shown by [Emb 01] that

$$\lambda_L = \lim_{u \to 0} \left(P[U_1 < u | U_2 = u] + P[U_2 < u | U_1 = u] \right),$$

and for distributions with infinite left endpoint such as the Gaussian or Gaussian Mixture this can be written

$$\lambda_L = \lim_{u \to 0} \left(P[X_1 < F_1^{-1}(u) | X_2 = F_2^{-1}(u)] + P[X_2 < F_2^{-1}(u) | X_1 = F_1^{-1}(u)] \right). \quad (3.8)$$

Let us consider the conditional GM distribution, which will enable us to compute the tail dependence coefficient. For the bivariate Gaussian distribution we the have standard result that, given parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$ and r, the distribution of X_2 conditional on $X_1 = x_1$ is

$$F(x_2 | X_1 = x_1) = \Phi(x_2; \mu_2 + r\frac{\sigma_2}{\sigma_1} x_1, \sigma_2^2(1 - r^2)).$$

How does this extend to the GM case? Applying some results from elementary probability theory, we have

$$F_{\text{GM}}(x_2|X_1 = x_1) = \sum_{k=1}^{M} P[(X_2 < x_2 \cap \text{state} = k)|X_1 = x_1]$$

$$= \sum_{k=1}^{M} P[X_2 < x_2|(X_1 = x_1 \cap \text{state} = k)]P[\text{state} = k|X_1 = x_1].$$

Now let

$$Q^{(k)}(x_1) = P[\text{state} = k|X_1 = x_1].$$

Applying Bayes' rule we have

$$Q^{(k)}(x_1) = \frac{P[X_1 = x_1|\text{state} = k]}{P[X_1 = x_1]} P[\text{state} = k].$$

We appear to be making things more and more complicated, but observe that by writing $x_2|X_1 = x_1$ as we have, we now know all of the probabilities that we require:

$$P[X_2 < x_2|(X_1 = x_1 \cap \text{state} = k)] = \Phi(x_2; \mu_2^{(k)} + \frac{\sigma_2^{(k)}}{\sigma_1^{(k)}}\rho_{12}^{(k)}x_1, \sigma_2^{(k)^2}(1 - \rho_{12}^{(k)^2})),$$

$$\frac{P[X_1 = x_1|\text{state} = k]}{P[X_1 = x_1]} = \frac{\phi(x_1; \mu_1^{(k)}, \sigma_1^{(k)})}{f_1(x_1)} \quad \text{and} \quad P[\text{state} = k] = p^{(k)}.$$

We now have that

$$Q^{(k)}(x_1) = \frac{\phi(x_1; \mu_1^{(k)}, \sigma_1^{(k)})}{f_1(x_1)} p^{(k)}$$

and

$$F(x_2|X_1 = x_1) = \sum_{k=1}^{M} Q^{(k)}(x_1)\Phi(x_2; \mu_2^{(k)} + \frac{\sigma_2^{(k)}}{\sigma_1^{(k)}}\rho_{12}^{(k)}x_1, \sigma_2^{(k)}(1 - \rho_{12}^{(k)^2})) \tag{3.9}$$

Now let us return to tail dependence. Denoting $x_1 = F_1^{-1}(u)$ and $x_2 = F_2^{-1}(u)$ and substituting (3.9) into (3.8) we have

$$\lambda_L = \lim_{x_1, x_2 \to -\infty} \left[\sum_{k=1}^{M} Q^{(k)}(x_1)\Phi\left(x_2; \mu_2^{(k)} + \frac{\sigma_2^{(k)}}{\sigma_1^{(k)}}\rho_{12}^{(k)}x_1, \sigma_2^{(k)}(1 - \rho_{12}^{(k)^2})\right) \right. \tag{3.10}$$

$$\left. + \sum_{k=1}^{M} Q^{(k)}(x_2)\Phi\left(x_1; \mu_1^{(k)} + \frac{\sigma_1^{(k)}}{\sigma_2^{(k)}}\rho_{12}^{(k)}x_2, \sigma_1^{(k)}(1 - \rho_{12}^{(k)^2})\right) \right]$$

First we look at the first term in the above expression. Consider the asymptotic behaviour of $Q^{(k)}(x_1)$. We have that

$$Q^{(k)}(x_1) = \frac{\phi(x_1; \mu_1^{(k)}, \sigma_1^{(k)})p^{(k)}}{F_1(x_1)}$$

$$= \frac{p^{(k)} \exp\left(\frac{-(x-\mu_1^{(k)})^2}{2\sigma_1^{(k)}}\right)}{\sum_{l=1}^{M} p^{(l)} \exp\left(\frac{-(x_1-\mu_1^{(l)})^2}{2\sigma_1^{(l)}}\right)} = \frac{1}{\sum_{l=1}^{M} \frac{p^{(l)}}{p^{(k)}} \exp\left(\frac{-(x_1-\mu_1^{(l)})^2}{2\sigma_1^{(l)}} + \frac{(x_1-\mu_1^{(k)})^2}{2\sigma_1^{(k)}}\right)}$$

Now for the components that have $\sigma_1^{(k)} < \max(\sigma_1^{(k)})$ or $\mu_1^{(k^*)} > \min(\mu_1^{(k^*)})$ we find that

$$\lim_{x_1 \to -\infty} Q^{(k)}(x_1) = 0.$$

For the components with $\sigma_1^{(k)} = \max(\sigma_1^{(k)})$ and $\mu_1^{(k^*)} = \min(\mu_1^{(k*)})$, which we denote k_1^-, we have

$$\lim_{x_1 \to -\infty} Q^{(k)}(x_1) = \frac{p^{(k)}}{\sum_{k \in k_1^-} p^{(k)}},$$

and exactly analogous arguments apply for the $Q^{(k)}(x_2)$ terms in equation (3.10). Thus we have

$$\lambda_L = \lim_{x_1, x_2 \to -\infty} \left[\sum_{k \in k_1^-} \frac{p^{(k)}}{\sum_{k \in k_1^-} p^{(k)}}(x_1) \Phi\left(x_2; \mu_2^{(k)} + \frac{\sigma_2^{(k)}}{\sigma_1^{(k)}} \rho_{12}^{(k)} x_1, \sigma_2^{(k)}(1 - \rho_{12}^{(k)^2})\right) \right. \quad (3.11)$$

$$\left. + \sum_{k \in k_2^-} \frac{p^{(k)}}{\sum_{k \in k_2^-} p^{(k)}} \Phi\left(x_1; \mu_1^{(k)} + \frac{\sigma_1^{(k)}}{\sigma_2^{(k)}} \rho_{12}^{(k)} x_2, \sigma_1^{(k)}(1 - \rho_{12}^{(k)^2})\right) \right]$$

Now observe that

$$\lim_{x_1, x_2 \to -\infty} \left[\Phi\left(x_2; \mu_2^{(k^*)} + \frac{\sigma_1^{(k^*)}}{\sigma_2^{(k^*)}} \rho_{12}^{(k^*)} x_1, \sigma_2^{(k^*)^2}(1 - \rho_{12}^{(k^*)^2})\right) \right] = \frac{1}{2}$$

if and only if $r = 1$, $\sigma_2^{(k)} = \max(\sigma_2^{(k)})$ and $\mu_2^{(k^*)} = \min(\mu_2^{(k^*)})$, otherwise we get

$$\lim_{x_1, x_2 \to -\infty} \left[\Phi\left(x_2; \mu_2^{(k^*)} + \frac{\sigma_1^{(k^*)}}{\sigma_2^{(k^*)}} \rho_{12}^{(k^*)} x_1, \sigma_2^{(k^*)^2}(1 - \rho_{12}^{(k^*)^2})\right) \right] = 0.$$

An analogous result holds for $\Phi\left(x_1; \mu_1^{(k)} + \frac{\sigma_2^{(k)}}{\sigma_1^{(k)}} r^{(k)} x_2, \sigma_1^{(k)^2}(1 - r^{(k)^2})\right)$. Substituting into equation (3.11) we have

$$\lambda_L = \sum_{k \in k^-} \frac{p^{(k)}}{\sum_{k \in k^-} p^{(k)}} \frac{1}{2} \mathbf{1}_{[r^{(k)}=1]} + \sum_{k \in k^-} \frac{p^{(k)}}{\sum_{k \in k^-} p^{(k)}} \frac{1}{2} \mathbf{1}_{[r^{(k)}=1]}$$

$$= \frac{\sum_{k^-} p^{(k \in k^-)} \mathbf{1}_{[r^{(k \in k^-)}=1]}}{\sum_{k^-} p^{(k^-)}}$$

The result for λ_U can be proved in a very similar manner, for brevity we omit it here.

$\diamondsuit$

Clearly a GM copula with $r^{(k)} \neq 1$ for all k has zero tail dependence. Observe, however, that we can simply add another distribution to our mixture with $\sigma^{(M+1)} > \max(\sigma^{(1)} \dots \sigma^{(M)})$ and $p^{(M+1)} > 0$, to yield $\lambda_L = \lambda_U = 1$. If we let $p^{(M+1)} \to 0^+$ then the impact of this additional distribution on the behaviour of the copula is negligible, but the tail dependence is still 1. Of course, we do not suggest that this is a practical approach to induce tail dependence into a copula. Rather, this serves to reinforce the point that tail dependence, since it is defined asymptotically, is of very little practical usefulness in working with copulas.

We have argued that the quantile dependence is of much more interest. As usual this can be computed pairwise as

$$\nu_L(u) = \frac{C_{GM}(u,u)}{u} \qquad \text{and} \qquad \nu_U(u) = \frac{\tilde{C}_{GM}(u,u)}{1-u},$$

which can easily be computed using the bivariate normal distribution function and the univariate normal inverse distribution function. The lower and upper quantile dependence for this copula together with the empirical sample are plotted in Figure 3.8. We have chosen the parameters so that

$$\nu_L(0.05) = \bar{\nu}_L(0.05) \qquad \nu_U(0.05) = \bar{\nu}_U(0.05) \quad \text{and} \quad \tau = \bar{\tau}.$$

Notice how the behaviour of ν_L and ν_U is close to that seen for the empirical data in Figure 2.2.

Thus the key dependence measures – Kendall's tau, Spearman's rho, tail dependence and quantile dependence – for the GM copula can be computed quickly and accurately given any parameter set. Next we suggest how to use these results to calibrate the GM copula to empirical data.

3.2.3 Implementation

There are many potential approaches, the following is just one example. While a maximum likelihood approach may be possible in theory the large number of parameters plus the problem of poor quality data mean that this is unlikely to be practical. Instead we would like something analogous to a moment based approach. We shall exploit the pairwise nature of parameters in the Gaussian Mixture model to calibrate to Kendall's tau, lower quantile dependence and upper quantile dependence.

To calibrate our model we first measure these empirically, with the lower and upper quantile dependence at a confidence level of 5% and 95% respectively, to give $\bar{T}_{ij}$, $\bar{\lambda}_{Lij}$

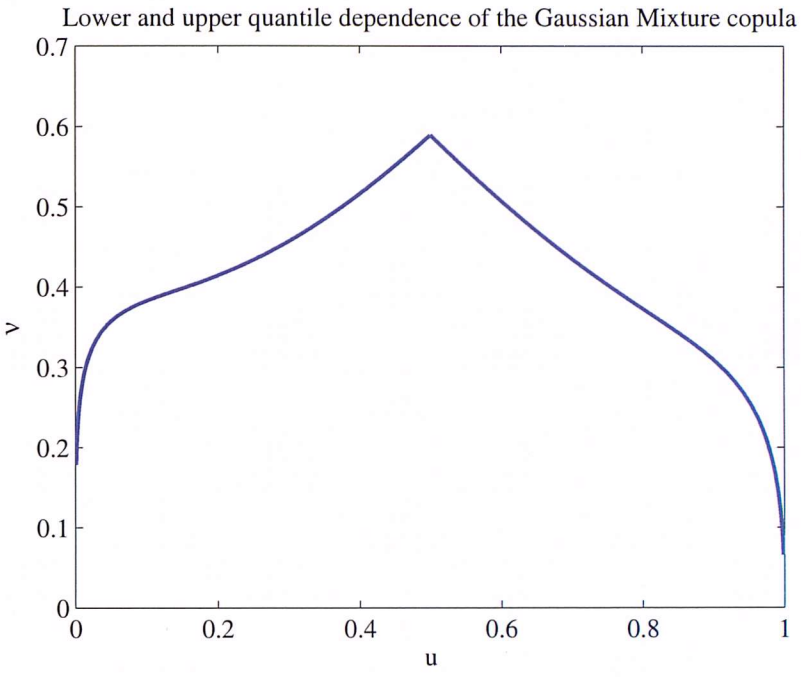


Figure 3.8: Quantiles of the Gaussian Mixture copula

and $\bar{\lambda}_{Uij}$. We therefore have $\frac{3}{2}d(d-1)$ empirical measurements to calibrate to. We would like to set up a copula with the same number of unknowns, in order to give the potential for a unique calibration. Let us assume than that we have a mixture of three normals, having means

$$\mu_1^{(1)} = \mu_2^{(1)} = 0, \qquad \mu_1^{(2)} = \mu_2^{(2)} = -1 \quad \text{and} \quad \mu_1^{(3)} = \mu_2^{(3)} = 1;$$

standard deviations

$$\sigma_1^{(1)} = \sigma_2^{(1)} = 1 \qquad \sigma_1^{(2)} = \sigma_2^{(2)} = 1 \quad \text{and} \quad \sigma_1^{(3)} = \sigma_2^{(3)} = 1;$$

and probabilities

$$p^{(1)} = 0.6 \qquad p^{(2)} = 0.2 \qquad p^{(3)} = 1 - p^{(1)} - p^{(2)} = 0.2.$$

The correlation matrices $R^{(1)}, R^{(2)}$ and $R^{(3)}$ are left unspecified, to yield the desired $\frac{3}{2}d(d-1)$ free parameters. Roughly speaking, for the ijth pair of assets, the correlation $R_{ij}^{(2)}$ controls the lower quantile dependence, $R_{ij}^{(3)}$ controls the upper quantile dependence, and all three correlations influence Kendall's tau. Now it remains to calibrate these correlations on a pairwise basis. From the expressions in the previous sections we have that

$$\nu_L(R^{(1)}, R^{(2)}, R^{(3)}) = \frac{1}{0.05} C(.05, .05; R^{(1)}, R^{(2)}, R^{(3)})$$

$$\nu_U(R^{(1)}, R^{(2)}, R^{(3)}) = \frac{1}{0.05}\left(1 - 0.95 - 0.95 + C(.95, .95; R^{(1)}, R^{(2)}, R^{(3)})\right)$$

and

$$\tau(R^{(1)}, R^{(2)}, R^{(3)}) = \tau(X_1, X_2; R^{(1)}, R^{(2)}, R^{(3)}),$$

where we have suppressed the dependence on the constant mean, variance and probability parameters. We can now define an objective function which measures the error in our calibration

$$E(R^{(1)}, R^{(2)}, R^{(3)}; \hat{\tau}, \hat{\nu}_L, \hat{\nu}_U) = \sqrt{(\tau - \hat{\tau})^2 + (\nu_L - \hat{\nu}_L)^2 + (\nu_U - \hat{\nu}_U)^2}.$$

We then simply minimise E with respect to $\{R^{(1)}, R^{(2)}, R^{(3)}\}$ separately for each pair of assets. It will not always be possible to find $\{R^{(1)}, R^{(2)}, R^{(3)}\}$ such that $E = 0$, but for a 'typical' pair of financial time series an exact calibration can normally be found using the values of μ, σ and p described above. For the SG-TH data we find that

$$r^{(1)} = -0.21 \qquad r^{(2)} = 0.67 \qquad r^{(3)} = 0.44.$$

Notice that we have calibrated pairwise, which is desirable for the reasons given in section §2.3.1.

One potential hitch is that some or all of the correlation matrices implied by the calibration may not be positive definite, but there exist methods for working on a correlation matrix to ensure positive definiteness (e.g. [Reb 99]). Also, we have chosen the values for μ, σ and p *a priori*, though experiments with real data seem to indicate that they work well. The point is that even with an *a priori* choice of some parameters, we still have sufficient flexibility to model the empirical dependence successfully.

Simulation from the Gaussian mixture distribution is straightforward, we are required only to simulate from multivariate normal distributions conditional on the realised value of a uniform state variable. We can transform these to a copula sample by the application of the univariate Gaussian mixture cdf, which is again simple since it is just a sum of Gaussian cdfs.

Gaussian Mixture Simulation Algorithm

- Compute the Cholesky decomposition $A^{(k)}$ of each dispersion matrix $R^{(k)}$
- Simulate d independent $N(0,1)$ random variables, $z_1, \ldots, z_d$
- Simulate a uniform random variate v independent of $z_1, \ldots, z_d$
- Set $l = k$ such that $\sum_{k=1}^{l} p^{(k)} < v$ and $\sum_{k=1}^{l+1} p^{(k)} > v$
- Set $x_i = (A_{i1}^{(l)} z_1 + \cdots + A_{ii}^{(l)} z_i)\sigma_i^{(l)} + \mu_i^l$ for $i = 1, \cdots, d$
- Set $u_i = F_{\mathrm{GM}_i}(x_i)$

Chapter 4

Non-Parametric Copula Models

We now consider a rather different approach to copula estimation: rather than specifying some parametric form *a priori*, we shall adopt a data driven approach to the construction of the copula function.

4.1 Empirical Sub-Copula

The most elementary form of non-parametric dependence is that introduced by [Deh 80]:

Definition 21 *Suppose we have a d-dimensional empirical sample of length, T, denoted $\left\{\bar{x}_1^{(t)}, \ldots, \bar{x}_d^{(t)}\right\}_{t=1}^{T}$, and having rank statistics $\left\{\bar{r}_1^{(t)}, \ldots, \bar{r}_d^{(t)}\right\}_{t=1}^{T}$. The function $\hat{C}'$ that is defined on the lattice*

$$\mathbf{L}_d^T = \left\{\left(\frac{i_1}{T}, \ldots, \frac{i_d}{T}\right) : i_j = 0, \ldots, T; j = 1, \ldots, d\right\}$$

as

$$\hat{C}'\left(\frac{i_1}{T}, \ldots, \frac{i_d}{T}\right) = \frac{1}{T}\sum_{t=1}^{T}\prod_{j=1}^{d}\mathbf{1}_{[\bar{r}_j^{(t)} \leq i_j]} \tag{4.1}$$

is called an empirical sub-copula.

This is widely referred to in the literature (see [Nel 99] and [Che 04]) as the empirical copula, but since it is not defined away from $\mathbf{L}_d^T$, we regard the term 'empirical sub-copula' to be correct.

Some authors (see, [Fer 04], [Sca 05]) have extended this definition to the whole of $[0,1]^d$ by setting

$$\hat{C}(u_1, \ldots, u_d) = \frac{1}{T}\sum_{t=1}^{T}\prod_{j=1}^{d}\mathbf{1}_{[\bar{r}_j^{(t)} \leq u_j T]}. \tag{4.2}$$

With this definition we have that the density[1] of the empirical copula is

$$\hat{c}(u_1, \ldots, u_d) = \frac{1}{T} \sum_{t=1}^{T} \prod_{j=1}^{d} \delta(u_j T - \bar{r}_j^{(t)})$$

where $\delta(x)$ is the Dirac delta function. This density is plotted in Figure 4.1 (where we represent the delta function as a vertical line at $x = 0$) – clearly this function is highly discontinuous. This makes it difficult to compare with parametric densities, and if we attempt to generate a sample from this copula we just generate random draws from the original distribution – in effect performing statistical 'bootstrapping', see [Efr 79]. Moreover, the unsmoothed estimator (4.2) has the drawback that, away from the lattice $\mathbf{L}_d^T$, $\hat{C}$ is not actually a copula, since it fails to satisfy the second requirement of Definition 1. To see this, consider the point $(1, \ldots, u_i, , 1, \ldots, 1)$ where $0 < u_i < \frac{1}{T}$. Since $\min_t(r_i^{(t)}) = 1$ we have that

$$\hat{C}(1, \ldots, 1, u_i, 1, \ldots, 1) = 0 \neq u_i.$$

Alternatively we can observe that $\hat{C}$ is obviously not uniformly continuous, which violates the continuity property (2.1).

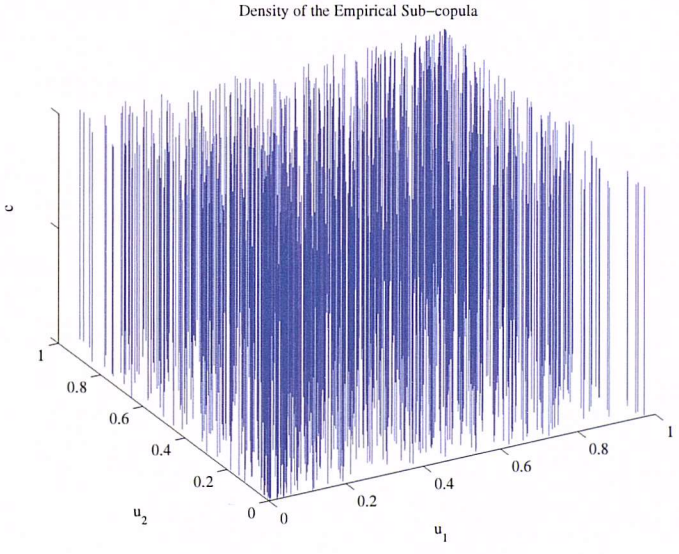


Figure 4.1: Density of the unsmoothed empirical sub-copula for SG-TW data

We would like to extend the empirical sub-copula to be defined over $[0, 1]^d$ in a way that: satisfies the definition of a copula; enables us to compare the empirical density to

[1]We use the term loosely – since the distribution function (4.2) is not absolutely continuous it does not posses a density in the strict sense.

that of some parametric copula; and allows us to generate Monte Carlo samples from outside of the original data. The former can be achieved using a Partition of Unity.

4.2 Partitions of Unity

Our first approach follows from the results of [Li 97], [Li 98] and [Kul 99],

Definition 22 *A collection of functions* $\phi_1, \ldots, \phi_n \in L^1([0,1])$ *is called a* Partition of Unity *if*

1. $\phi_i(u) \geq 0$ *for* $i = 1, \ldots, n$
2. $\int_0^1 \phi_i(t)dt = \frac{1}{n}$ *for* $i = 1, \ldots, n$
3. $\sum_{i=1}^n \phi(u) = 1$ *for all* $u \in [0,1]$

We also denote by $\Delta_{i_1,\ldots,i_d}(\cdot)$ the d-dimensional partial finite difference operator, so that, for example, given a 2-dimensional copula C we have

$$\Delta_{i_1,i_2}(C) = C\left(\frac{i_1}{n}, \frac{i_2}{n}\right) - C\left(\frac{i_1-1}{n}, \frac{i_2}{n}\right) - C\left(\frac{i_1}{n}, \frac{i_2-1}{n}\right) + C\left(\frac{i_1-1}{n}, \frac{i_2-1}{n}\right).$$

We can now use Partitions of Unity to approximate copulas via the following theorem:

Theorem 11 Partition of Unity Extension *Given a Partition of Unity* $\phi_1, \ldots, \phi_n$ *and a* d*-dimensional sub-copula* C' *defined on the lattice* $\mathbf{L}_d^n$*, the function* $C_\phi : [0,1]^d \to [0,1]$ *defined by*

$$C_\phi(\mathbf{u}; C') = n^d \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} \Delta_{i_1 \ldots i_d}(C') \prod_{j=1}^{d} \int_0^{u_j} \phi_{i_j}(t)dt$$

is a copula.

Proof The 2-dimensional version of this result for C' a copula is proved by [Li 97], who use the equivalence between 2-dimensional copulas and Markov operators demonstrated by [Dar 92]. We use the defining properties of the copula directly to construct the more general proof:

1. C_ϕ is grounded. This follows directly:

$$C_\phi(u_1, \ldots, u_{k-1}, 0, u_{k+1}, \ldots, u_d)$$
$$= n^d \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} \Delta_{i_1 \ldots i_d}(C') \prod_{j \neq k}^{d} \int_0^{u_j} \phi_{i_j}(t)dt \times \int_0^0 \phi_{i_k}(t)dt = 0.$$

2. C_ϕ has uniform marginals. This follows from the properties of C' and $\phi_1, \ldots, \phi_n$:

$$
\begin{aligned}
& C_\phi(1, \ldots, 1, u_k, 1, \ldots, 1) \\
&= n^d \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} \Delta_{i_1 \ldots i_d}(C') \prod_{j \neq k}^{d} \int_0^1 \phi_{i_j}(t)dt \times \int_0^{u_k} \phi_{i_k}(t)dt \\
&= n \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} \Delta_{i_1 \ldots i_d}(C') \int_0^{u_k} \phi_{i_k}(t)dt \\
&= n \sum_{i_k=1}^{n} \Delta_{n,\ldots,n,i_k,n,\ldots,n}(C') \int_0^{u_k} \phi_{i_k}(t)dt \\
&= n \sum_{i_k=1}^{n} \frac{1}{n} \int_0^{u_k} \phi_{i_k}(t)dt \\
&= \int_0^{u_k} [\phi_{i_1}(t) + \cdots + \phi_{i_n}(t)]dt \\
&= \int_0^{u_k} dt = u_k.
\end{aligned}
$$

3. C_ϕ is d-increasing. To see this consider the density

$$
c_\phi(\mathbf{u}) = \frac{\partial^d}{\partial u_1 \ldots \partial u_d} C_\phi(\mathbf{u}) = n^d \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} \Delta_{i_1 \ldots i_d}(C') \prod_{j=1}^{d} \phi_{i_j}(u_j).
$$

We have that $\Delta_{i_1 \ldots i_d}(C') \geq 0$ since C' is a sub-copula (and hence d-increasing), and we have that $\phi_i \geq 0$ from the definition. Thus

$$
c_\phi(\mathbf{u}) = \frac{\partial^d}{\partial u_1 \ldots \partial u_d} C_\phi(\mathbf{u}) \geq 0 \ \forall \ \mathbf{u} \in [0,1]^d.
$$

Since the partial derivatives of C_ϕ are everywhere non-negative, we have that C_ϕ is everywhere d-increasing.

Hence C_ϕ always satisfies the requirements to be a copula function.

$\diamondsuit$

This theorem could be generalized slightly by having a different order of partition in each of the d dimensions, and we need not use the same partition of unity in each dimension. To ease notation we have stated and proved the slightly simplified case, but the above result can easily be extended.

Corollary 12 *Given a d-dimensional sample of length T and a partition of unity $\phi_1, \ldots, \phi_n$, the function*

$$C_\phi(\mathbf{u}; \hat{C}') = n^d \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} \Delta_{i_1 \ldots i_d}(\hat{C}') \prod_{j=1}^{d} \int_0^{u_j} \phi_{i_j}(t) dt$$

is a copula if and only if $T = kn$ for some positive integer k.

Proof The previous proof required that C' satisfy the conditions for a sub-copula on the lattice $\mathbf{L}_n^d$. Given a sample of length T, we can form an empirical sub-copula $\hat{C}'$ on $\mathbf{L}_T^d$ using (4.1). So $\hat{C}'$ is defined on $\mathbf{L}_n^d$ if and only if $T = kn$ for integer k.

The requirement that $T = kn$ for integer k has been overlooked in the literature. The empirical copulas defined by [Dur 00b] and [Che 04] §5.5 implicitly use $n = T$, which considerably limits the flexibility of this approximation (see §3.3.4 for a discussion of how the choice of n affects the approximation). The 'Bernstein copula' defined by [San 04] §4 will in general fail the uniformity of marginals requirement, and hence fail to be a copula. This may seem like an academic point, but would, for example, cause difficulties in using samples from the 'copula' to simulate from a marginal distribution. Of course, the requirement that $T = kn$ for integer k is rather restrictive, not least if T happens to be prime. We resolve this problem by defining the empirical copula to be the extension of the empirical sub-copula to $[0,1]^d$, using a simple Partition of Unity with $n = T$. Consider the set of functions

$$\chi_{i,n}(t) = \begin{cases} 0 & \text{if} \quad t < \frac{i-1}{n} \\ 1 & \text{if} \quad \frac{i-1}{n} \leq t < \frac{i}{n} \\ 0 & \text{if} \quad t \geq \frac{i}{n} \end{cases} \qquad \text{for} \quad i = 1, \ldots, n.$$

Clearly $\chi_{i,n}(t)$ is a Partition of Unity, which we can use to extend the empirical sub-copula to the unit hypercube.

Definition 23 *Suppose we have a d-dimensional empirical sample of length T, which has associated with it an empirical sub-copula $\hat{C}'$. The function $\hat{C} : [0,1]^d \to [0,1]$ defined as*

$$\hat{C}(\mathbf{u}; \hat{C}') = T^d \sum_{i_1=1}^{T} \cdots \sum_{i_d=1}^{T} \Delta_{i_1 \ldots i_d}(\hat{C}') \prod_{j=1}^{d} \int_0^{u_j} \chi_{i_j,T}(t) dt \tag{4.3}$$

is called the empirical copula. *The function $\hat{c} : [0,1]^d \to [0,\infty)$ defined as*

$$\hat{c}(\mathbf{u}; \hat{C}') = T^d \sum_{i_1=1}^{T} \cdots \sum_{i_d=1}^{T} \Delta_{i_1 \ldots i_d}(\hat{C}') \prod_{j=1}^{d} \chi_{i_j,T}(u_j) \tag{4.4}$$

is called the empirical copula density.

We remark that (4.4) is precisely the 'empirical dependence function' defined by [Deh 80].

Thus we have an extensions of the empirical sub-copula that satisfies the definition of a copula, but we still have the problem that the estimated density is heavily dependent on the fine structure of the sample. This can be solved by smoothing the empirical copula by applying a Partition of Unity with $n \ll T$ to $\hat{C}$ itself. Since $\hat{C}$ is defined on the whole of $[0,1]^d$, we are not limited to choosing $n = \frac{T}{k}$, as we would be had we tried to smooth $\hat{C}'$ directly. We show how to choose n in §4.4.1, but first we look at some specific examples of $\phi_i(t)$.

4.2.1 Checkerboard Approximation

Perhaps the simplest partition of unity is the one we used to define the empirical sub-copula, which is referred to as the checkerboard approximation. This arises when we set

$$C_\chi(\mathbf{u}) = n^d \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} \Delta_{i_1 \ldots i_d}(C') \prod_{j=1}^{d} \int_0^{u_j} \chi_{i_j}(t)dt$$

Clearly this represents a very simple approach to estimation, but it has been shown to converge in a uniform and in a strong sense [Li 98]. If we apply a d-dimensional checkerboard approximation to the empirical copula (4.3) we get a smoothed estimate of the density

$$c_\chi(\mathbf{u}) = n^d \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} \Delta_{i_1,\ldots,i_d}(\hat{C}) \prod_{j=1}^{d} \chi_{i_j,n}(u_j) \tag{4.5}$$

In fact we can see that for this choice of partition of unity we just have the histogram estimator of order n, which we plot using the SG-TW data in Figure 4.2. This link already suggests a relationship between partitions of unity and density estimation which we will explore more fully in what follows.

[Dur 00a] derived the following expressions for the Kendall's tau and Spearman's rho of the Checkerboard approximation:

$$\tau(C_\chi) = 1 - \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{p=1}^{n}\sum_{q=1}^{n} \Delta_{ij}(C')\Delta_{pq}(C')\xi_{ip}\xi_{qj}$$

and

$$\rho_S(C_\chi) = 3\sum_{i=1}^{n}\sum_{j=1}^{n} \Delta_{ij}(C')\omega_{ij} - 3,$$

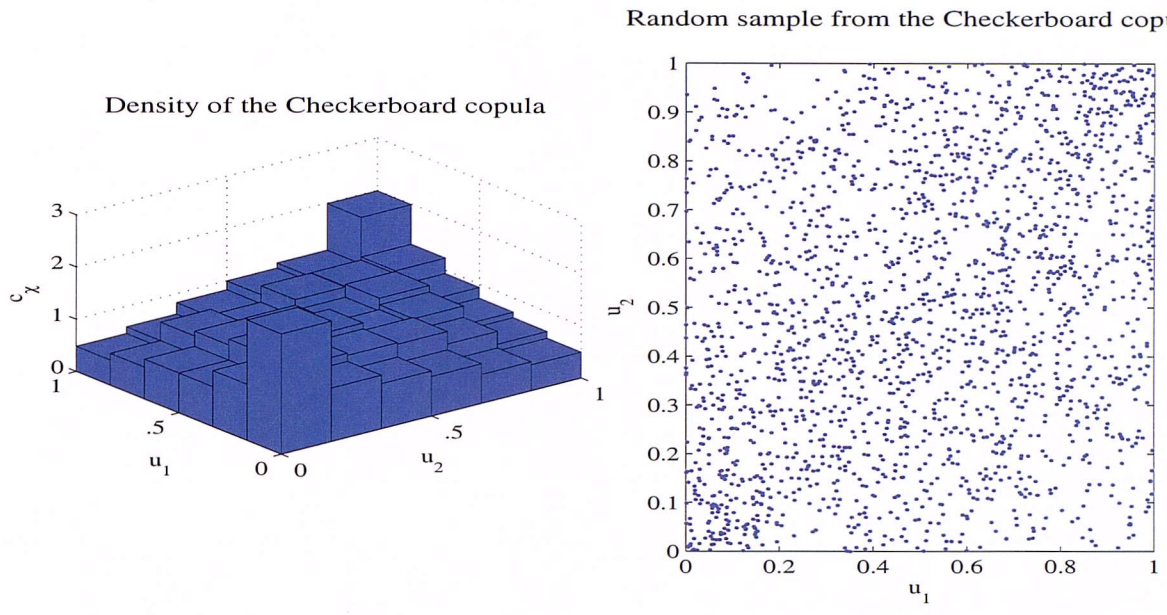


Figure 4.2: Density of the Checkerboard approximation with $n = 6$ for the sample data

where

$$\xi_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 2 & \text{if} \quad i > j \\ 0 & \text{if} \quad i < j \end{cases} \qquad \text{and} \qquad \omega_{ij} = \frac{(2n - 2i + 1)(2n - 2j + 1)}{n^2}.$$

Since the main use of dependence measures is in calibration - and we do not need to calibrate the empirical copulas is the usual sense - these expressions are of largely academic interest. [Dur 00a] show that when we use C_χ to approximate a parametric copula, $\tau(C_\chi)$ and $\rho_S(C_\chi)$ converge from below as $n \to \infty$. These authors also show that

$$\lambda_L(C_\chi) = \lambda_U(C_\chi) = 0,$$

i.e. the Checkerboard approximation has zero tail dependence, which they regard as "a real problem in financial applications". We have previously argued that this is not in fact the case.

4.2.2 Bernstein Approximation

We have seen how one simple partition of unity can be applied to approximate the empirical copula, but we would like to find one with better properties, especially regarding the smoothness of the copula density. An improvement can be made using the Bernstein

polynomials

$$B_{i,n}(t) = \binom{n}{i} t^i (1-t)^{n-i}, \tag{4.6}$$

which we plot for $n = 10$ in Figure 4.3.

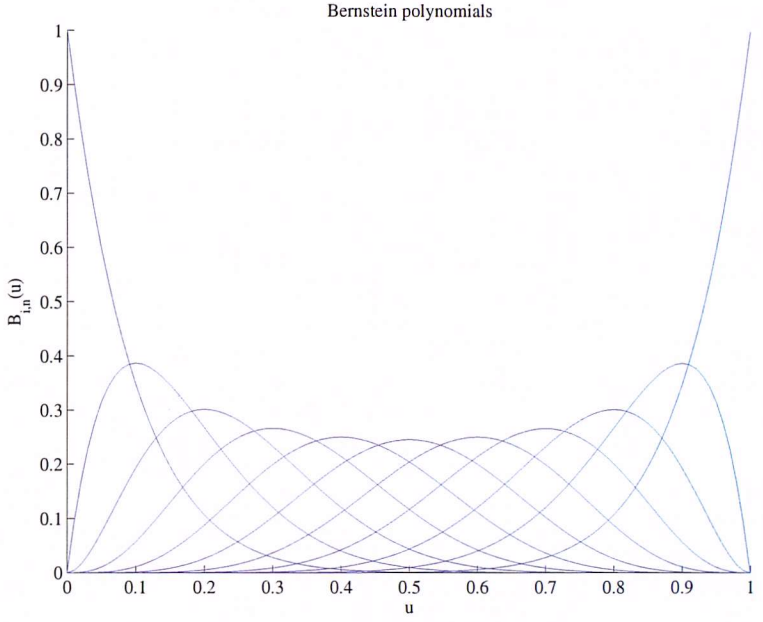


Figure 4.3: The Bernstein polynomial partition of unity for $n = 10$

The name comes from Bernstein's theorem which states that any continuous function $f : [0,1] \to (-\infty, \infty)$ can be approximated by a sequence of polynomials as

$$\mathbf{B}_n(f)(x) = \sum_{i=0}^{n} f\left(\frac{i}{n}\right) B_{i,n}(x),$$

first used by [Ber 12] to give a simple proof of the Weierstraß approximation theorem, and extended to the multivariate case by [But 53] and to one-dimensional density estimation by [Vit 75]. Since it can be shown that the Bernstein polynomials (4.6) satisfy the conditions to be a partition of unity, we can set

$$\phi_i(t) = B_{i-1,n-1}(t)$$

to construct the Bernstein approximation

$$C_B(\mathbf{u}) = n^d \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} \Delta_{i_1,\ldots,i_d}(C') \prod_{j=1}^{n} \int_0^{u_j} B_{i_j-1,n-1}(t)dt. \tag{4.7}$$

Since (see [Bab 02]),

$$\sum_{k=i}^{n} B_{k,n}(t) = n \int_0^t B_{i-1,n-1}(x)dx;$$

and (from the definition of the partial difference operator)

$$\sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} \Delta_{i_1,\ldots,i_d}(C') \prod_{j=1}^{n} \sum_{k=i_j}^{n} B_{k,n}(u_j)$$

$$= \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} C'\left(\frac{i_1}{n}, \ldots, \frac{i_d}{n}\right) \prod_{j=1}^{n} B_{i_j,n}(u_j);$$

and (by definition)

$$C'(\ldots, 0, \ldots) = 0;$$

we have that:

$$C_B(\mathbf{u}) = n^d \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} \Delta_{i_1,\ldots,i_d}(C') \prod_{j=1}^{n} \int_0^{u_j} B_{i_j-1,n-1}(t)dt$$

$$= \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} \Delta_{i_1,\ldots,i_d}(C') \prod_{j=1}^{n} \sum_{k=i_j}^{n} B_{k,n}(u_j)$$

$$= \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} C'\left(\frac{i_1}{n}, \ldots, \frac{i_d}{n}\right) \prod_{j=1}^{n} B_{i_j,n}(u_j)$$

$$= \sum_{i_1=0}^{n} \cdots \sum_{i_d=0}^{n} C'\left(\frac{i_1}{n}, \ldots, \frac{i_d}{n}\right) \prod_{j=1}^{n} B_{i_j,n}(u_j) = \mathbf{B}_n(C')(\mathbf{u}).$$

In other words, using Bernstein polynomials as the partition of unity to extend a sub-copula is equivalent to taking the Bernstein approximation of the sub-copula itself. Differentiating (4.7) we have that the density of the Bernstein approximation is just

$$c_B(\mathbf{u}) = \frac{\partial^d}{\partial u_1 \cdots \partial u_d} \mathbf{B}_n(C')(\mathbf{u}) = \frac{\partial^d}{\partial u_1 \cdots \partial u_d} C_B(\mathbf{u}) \tag{4.8}$$

$$= n^d \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} \Delta_{i_1,\ldots,i_d}(C') \prod_{j=1}^{d} B_{i_j-1,n-1}(u_j)$$

We have that this approximation converges strongly to the true copula C thanks to the results of [Li 98] and [Kul 99], while [San 05] have studied the L_2 properties of this approximation and outline some applications to portfolio optimisation. Using the above approximation on the empirical copula defined in 4.3, we have

$$c_B(\mathbf{u}) = n^d \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} \Delta_{i_1,\ldots,i_d}(\hat{C}) \prod_{j=1}^{d} B_{i_j-1,n-1}(u_j),$$

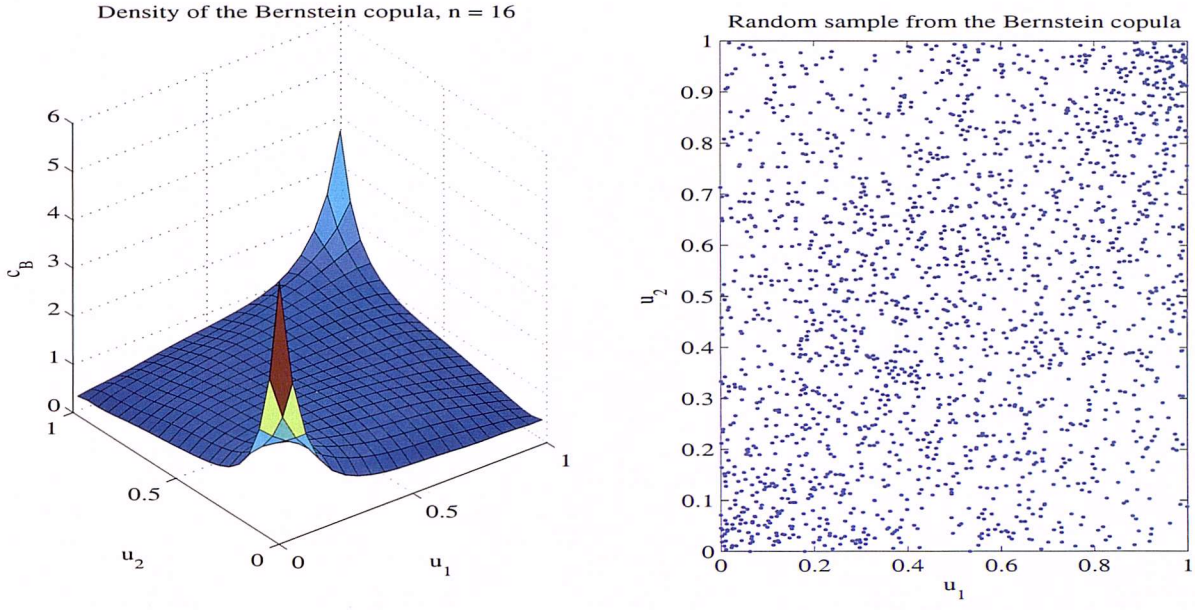


Figure 4.4: Bernstein approximation applied to the sample data with $n = 16$

which we plot for $n = 16$ in Figure 4.4.

Recall the form of the beta distribution which has density function $\beta_{p,q}(u) : [0,1] \to [0,\infty)$ defined as

$$\beta_{p,q}(u) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}(1-u)^{q-1}x^{p-1}.$$

We observe that the Bernstein coefficient multiplied with $(n+1)$ is identical to the beta density function with $p = n+1$ and $q = n-i+1$, i.e.

$$(n+1)B_{i,n}(u) = (n+1)\begin{pmatrix} n \\ i \end{pmatrix} u^i(1-u)^{n-i}$$

$$= \frac{\Gamma(n+2)}{\Gamma(i+1)\Gamma(n-i+1)}u^i(1-u)^{n-i} = \beta_{i+1,n-i+1}(u)$$

for all $u \in [0,1]$. Hence we can write the Bernstein density estimator as a weighted sum of n multi-variate beta densities

$$c(u_1,\ldots,u_d) = \sum_{i_1=1}^{n}\cdots\sum_{i_d=1}^{n}\Delta_{i_1,\ldots,i_d}(C)\prod_{j=1}^{d}\beta_{i_j,n-i_j+1}(u_j).$$

This property will be very useful later on when we want to generate samples from the Bernstein approximation.

[Dur 00a] derived the following expressions for the Kendall's tau and Spearman's rho of the Bernstein approximation:

$$\tau(C_B) = 1 - \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{p=1}^{n}\sum_{q=1}^{n}\Delta_{ij}(C')\Delta_{pq}(C')\gamma_{ip}\gamma_{qj}$$

and

$$\rho_S(C_B) = \frac{12}{(n+1)^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_{ij}(C') - 3,$$

where

$$\gamma_{ij} = \frac{(i-j)\begin{pmatrix} n \\ i \end{pmatrix}\begin{pmatrix} n \\ j \end{pmatrix}}{(2n-i-j)\begin{pmatrix} 2n-1 \\ i+j-1 \end{pmatrix}}.$$

Again [Dur 00a] show that the tail dependence of this copula is always zero, again we suggest that this is not a significant problem.

4.3 Kernel Estimators

Another approach to smoothing non-parametric density functions is kernel estimation. We discuss the general case briefly, see [Sil 86] or [Wan 95] for further details.

Definition 24 *A function* $k_h : (-\infty, \infty) \to [0, \infty)$ *is called a* kernel *if it satisfies the conditions*

1. k_h *is normalised; i.e.* $\int k_h(t)dt = 1$
2. k_h *is non-negative; i.e.* $k_h(x) \geq 0$ *for all* x

A kernel function usually takes the form $k_h(x) = k\left(\frac{x}{h}\right)$, *where* $h \in (0, \infty)$ *is referred to as the* smoothing parameter.

Notice that 1. and 2. together are the necessary and sufficient conditions for k to be a density function. With a function of this type we can construct a non-parametric estimate of the density of an empirical sample, $\{\bar{x}^{(t)}\}_{t=1}^{T}$, as:

$$\hat{f}(x) = \frac{1}{Th}\sum_{t=1}^{T} k\left(\frac{x - \bar{x}^{(t)}}{h}\right);$$

and of the distribution function as:

$$\hat{F}(x) = \int_{-\infty}^{x} \hat{f}(t)dt.$$

Notice how the definition of the kernel function ensures that $\hat{f}(x)$ and $\hat{F}(x)$ respectively satisfy the conditions for a density and distribution function.

Typical choices of functional form for k include the Epanechnikov kernel:

$$k_{\mathrm{E}}(x) = \max\left(0, \frac{3}{4\sqrt{5}}(1 - 1/5x^2)\right),$$

which can be shown (see [Sil 86], Chap. 3) to be most efficient in a mean square error sense for densities defined on $(-\infty, \infty)$; and the Gaussian kernel

$$k_{\mathrm{G}}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) = \phi(x),$$

often chosen for its differentiability properties. We will focus on the latter since the Gaussian property will allow us to make the link between this non-parametric estimator and the parametric GM copula we examined in the previous chapter.

The value of the smoothing parameter affects the performance of the kernel estimator more than the particular choice of functional form. The goal is usually to choose h to minimise the integrated mean square error relative to the true density $f(x)$. This involves a compromise, since reducing h will generally reduce the bias while increasing the variance, see [Sil 86] for details. We shall look at a method for choosing h optimally in practice in §4.4.1.

The kernel approach extends quite naturally to the multivariate case, where use an estimator of the density of the form

$$\hat{f}(\mathbf{x}) = \frac{1}{T|H|} \sum_{t=1}^{T} k\left(\mathbf{x} - \bar{\mathbf{x}}^{(t)}; H\right),$$

where k is now a multi-variate kernel function such that

$$\int_{(-\infty,\infty)^d} k(\mathbf{t}; H)\mathrm{d}\mathbf{t} = 1 \quad \text{and} \quad k(\mathbf{t}; H) \geq 0 \quad \text{for all } \mathbf{t},$$

and H is a $d \times d$ matrix of positive smoothing parameters.

The simplest way to construct $k(\mathbf{t}; H)$ is to use products of univariate kernels k_i, so that we have

$$k(\mathbf{t}; h) = \prod_{i=1}^{d} k_i\left(t_i; H_{ii}\right),$$

where H is now a diagonal matrix. Using Gaussian kernels, the estimate of the density then becomes:

$$\hat{f}(\mathbf{x}; H) = \frac{1}{T|H|} \sum_{t=1}^{T} \prod_{i=1}^{d} \phi\left(\frac{x_i - \bar{x}_i^{(t)}}{H_{ii}}\right); \tag{4.9}$$

and the distribution function is estimated as:

$$\hat{F}(\mathbf{x};H) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} \hat{f}(\mathbf{x};H)\mathrm{d}\mathbf{x} = \frac{1}{T}\sum_{t=1}^{T}\prod_{i=1}^{d}\Phi\left(\frac{x_i - \bar{x}_i^{(t)}}{H_{ii}}\right).$$

[Sil 86] suggests an extension of this approach by using the multivariate Gaussian distribution to get a density estimate of the form

$$\hat{f}(\mathbf{x}) = \frac{1}{T|H|}\sum_{t=1}^{T}\Phi_d\left((\mathbf{x} - \bar{\mathbf{x}}^{(t)})^T; H\right),$$

where the smoothing matrix is set equal to some smoothing constant c multiplied by the covariance matrix of the sample data

$$H = c\hat{\Sigma}.$$

The idea being to choose a multivariate kernel function which is the same 'shape' as the sample data.

The most direct approach to smoothing of the copula density would be to apply (4.9) to the copula sample $\{\bar{u}_1^{(t)}, \ldots, \bar{u}_d^{(t)}\}_{t=1}^T$. Unfortunately, this approach will not work well in practice due to the phenomenon of boundary bias. We refer to [Wan 95] Chap. 5 for details, but the problem is essentially a result of the kernels assigning probability mass outside the domain of the true density. Clearly this is not a problem if the density takes non-zero value over $(-\infty, \infty)^d$, but for densities which are defined only on a finite interval the standard from of kernel estimator may not be optimal. This problem is particularly acute for copula densities, which are of course defined only on $[0,1]^d$ and often take maximal values at or near the boundary $\partial[0,1]^d$.

To circumvent this problem, [Fer 03] and [Ros 03] make use of the relation (2.2) and a Gaussian kernel to estimate the copula of a (not-necessarily uniform) sample $\{\bar{x}_1^{(t)}, \ldots, \bar{x}_d^{(t)}\}_{t=1}^T$ as

$$\bar{C}(\mathbf{u}) = \bar{F}(\bar{\zeta}) = \frac{1}{T}\sum_{t=1}^{T}\prod_{i=1}^{d}\Phi((\bar{\zeta}_i - \bar{x}_i^{(t)})/H_{ii})$$

where $\bar{\zeta} = (\bar{\zeta}_1, \ldots, \bar{\zeta}_d)'$ and $\bar{\zeta}_i = \inf_x\{x : \bar{F}_i(x;\bar{x}) \geq u_i\}$. Essentially this is just estimating the multivariate distribution via a kernel and normalising the marginals to estimate the copula indirectly. These authors show that this approach will converge for $H_{ii} \to 0$ as $T \to \infty$ and $TH_{ii}^2 \to 0$, and they use this estimator to analyse an empirical data set. In this approach the smoothed copula density is

$$\bar{c}(\mathbf{u}) = \frac{\bar{f}(\bar{\zeta})}{\prod_{i=1}^{d}\bar{f}_i(\bar{\zeta}_i)} = T^{d-1}\frac{\sum_{t=1}^{T}\prod_{i=1}^{d}\phi((\bar{\zeta}_i - \bar{x}_i^{(t)})/h_i)}{\prod_{i=1}^{d}\sum_{t=1}^{T}\phi((\bar{\zeta}_i - \bar{x}_i^{(t)})/h_i)}, \tag{4.10}$$

which can be seen to be a kernel estimate of (2.4). We can easily extend this model to include cross-terms in the fashion described above, so that we have

$$\bar{c}(\mathbf{u}) = T^{d-1} \frac{\sum_{t=1}^{T} \Phi_d \left((\bar{\zeta} - \bar{\mathbf{x}}^{(t)}); H \right)}{\prod_{i=1}^{d} \sum_{t=1}^{T} \phi((\bar{\zeta}_i - \bar{x}_i^{(t)}); H_{ii})}, \tag{4.11}$$

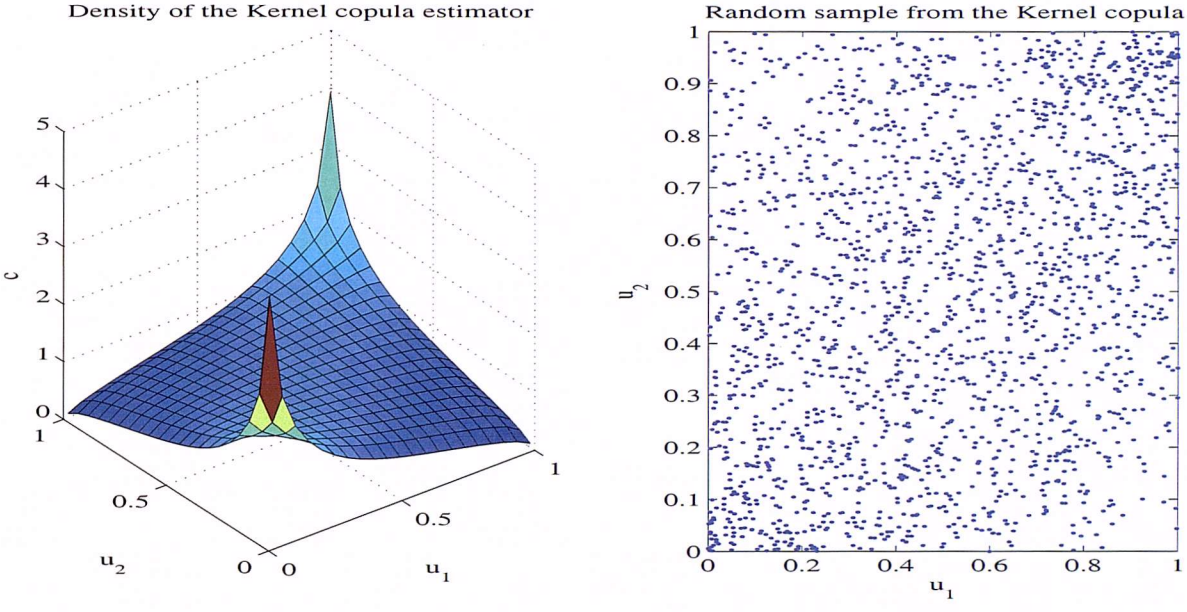


Figure 4.5: Density of the Kernel copula with cross terms included

We shall examine the practicalities of this approach in the next section, but we first observe that using a Gaussian kernel to approximate a copula is equivalent to using a Gaussian Mixture copula with T components, where

$$p^{(k)} = \frac{1}{T} \quad \mu^{(k)} = \bar{x}^{(k)} \quad \text{and} \quad \Sigma_{ij}^{(k)} = H_{ij} \quad \text{for } k = 1 \ldots, T.$$

We can thus use (3.4) and (3.6) to construct estimates for the Kendall's tau and Spearman's rho of the kernel approximation, analogous to those derived by [Dur 00a] for the Checkerboard and Bernstein copulas. We have that

$$\tau(\bar{C}) = \frac{2}{T^2} \sum_{k=1}^{T} \sum_{l=1}^{T} \Phi \left(\frac{\bar{x}_1^{(l)} - \bar{x}_1^{(k)}}{2H_{11}}, \frac{\bar{x}_2^{(l)} - \bar{x}_2^{(k)}}{2H_{22}}; \frac{H_{12}^2}{H_{11}H_{22}} \right)$$
$$+ \Phi \left(\frac{\bar{x}_1^{(k)} - \bar{x}_1^{(l)}}{2H_{11}}, \frac{\bar{x}_2^{(k)} - \bar{x}_2^{(l)}}{2H_{22}}; \frac{H_{12}^2}{H_{11}H_{22}} \right) - 1$$

and

$$\rho_S(\bar{C}) = \frac{6}{T^3} \sum_{k=1}^{T} \sum_{l=1}^{T} \sum_{m=1}^{T} \Phi \left(\frac{\bar{x}_1^{(l)} - \bar{x}_1^{(k)}}{2H_{11}}, \frac{\bar{x}_2^{(m)} - \bar{x}_2^{(k)}}{2H_{22}}; \frac{H_{12}}{H_{11}H_{22}} \right)$$

$$+\Phi\left(\frac{\bar{x}_1^{(k)}-\bar{x}_1^{(l)}}{2H_{11}},\frac{\bar{x}_2^{(k)}-\bar{x}_2^{(m)}}{2H_{22}};\frac{H_{12}}{H_{11}H_{22}}\right)-3.$$

We can also use the Gaussian Mixture interpretation to see that the upper and lower tail dependence coefficients of the kernel copula will be zero unless $\rho_{12}=1$, i.e. $H_{12}=H_{11}H_{22}$.

4.4 Implementation

4.4.1 Calibration

We now consider the question of 'calibrating' the nonparametric estimators. In a sense these estimators are automatically calibrated, since the empirical data is used directly in the construction of the copula. Indeed it is this property which is the main strength of the non-parametric approach – we avoid the unrealistic behaviour of the simple parametric models (e.g. the symmetry of elliptical copulas, homogeneity of Archimedean copulas) without the implementation problems created by some of the more complex parametric models. We do however need to select an optimal value for the smoothing parameter: n for the partition of unity estimators and H for the kernel estimator. As with other forms of non-parametric density estimation, this will involve a trade-off between minimising bias and minimising variance in the estimator.

One way to quantify the error in an estimator is the integrated mean square error (IMSE) between the true density $f(\mathbf{x})$ and our estimate $\bar{f}(\mathbf{x};h)$,

$$\text{IMSE}(\hat{f})=\int \mathbf{E}[\bar{f}(\mathbf{x})-f(\mathbf{x})]^2 d\mathbf{x}$$

$$=\int[\mathbf{E}[\hat{f}(\mathbf{x})]-f(\mathbf{x})]^2 d\mathbf{x}+\int \mathbf{E}[\hat{f}(\mathbf{x})^2]d\mathbf{x}\simeq\int \text{Bias}^2 d\mathbf{x}+\int \text{Variance}\, d\mathbf{x}$$

Computing IMSE is not straightforward since we do not know what $f(\mathbf{x})$ is. One approach suggested in the literature is to assume some functional form for $f(\mathbf{x})$ and thus derive an analytic formula for IMSE(h), which can then be minimised. If the true density is multivariate Gaussian, it can be shown (see [Sil 86], pp 85-87) that the optimal smoothing parameter is

$$h_i=\sigma_i T^{-1/5}, \tag{4.12}$$

where σ_i is the standard deviation of $\left\{\bar{x}_i^{(t)}\right\}_{t=1}^{T}$ – this is the approach used by [Fer 03]. Of course we do not actually believe that the true density really is multivariate Gaussian, we simply use this as a first approximation to choose h, and then let the non-parametric

copula uncover the true structure of the empirical distribution. We can also use (4.11), and set

$$H = \Sigma T^{-1/5} \tag{4.13}$$

where Σ denotes the covariance of the empirical data.

Trying to compute some analytical form for $IMSE(n)$ is not really feasible for the Partition of Unity approach, mainly because the copula density, which in this case we estimate directly, takes such an irregular form. Instead we shall use Likelihood Cross Validation, were we choose the value of n that maximises the quantity

$$\mathrm{LCV}(n) = \frac{1}{T}\sum_{t=1}^{T} \ln\left(\hat{f}_{-t}(\bar{\mathbf{x}}^{(t)}; n)\right).$$

where $\hat{f}_{-t}$ denotes the density estimate constructed using all of the data points except $\bar{\mathbf{x}}^{(t)}$. The idea here is that we remove one point from the empirical sample, construct our estimate of the density, then estimate the log-likelihood at the extra point. There is nothing special about the choice of point to remove, so we average over all points to construct a score function $LCV(n)$ – see [Sil 86] Chap. 3 for details. Implementation of this method is quite simple for the Bernstein approximation. $LCV(n)$ for the SG-TH data is shown in Figure 4.6, where s we find $n_{\mathrm{opt}} = 16$. We were not able to apply this methodology to the Checkerboard estimator, since here the density takes zero value on $[0,1]^d$ in non-trivial cases.

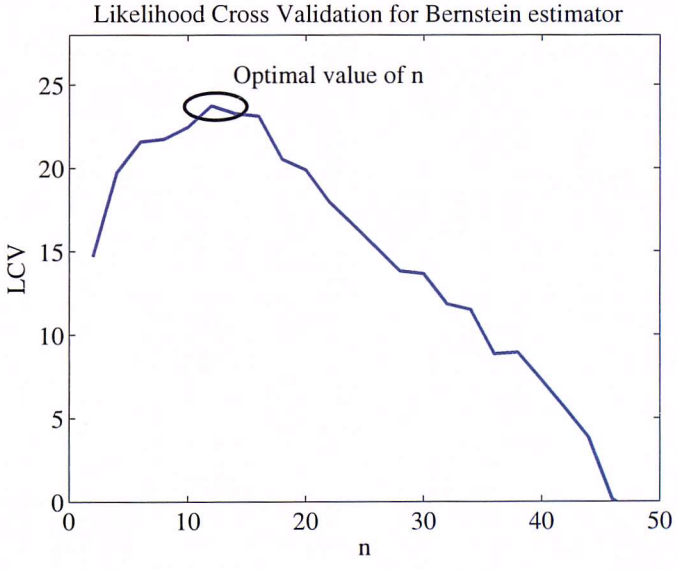


Figure 4.6: Likelihood Cross Validation score for the Bernstein approximation

We compare the non-parametric estimators by testing their ability to reproduce the (known) density of a parametric copula. We use the GM copula, which we suggest displays behaviour similar to empirical data from financial time series, with the sample parameters given in §3.2.3. We use the following algorithm to compute IMSE empirically.

IMSE Algorithm

- Compute the true density of the GM copula on the grid

$$\Big\{(u_1 \times u_2) = (0.01, \ldots, .99) \times (0.01, \ldots, .99)\Big\}$$

  denoted $c(\mathbf{u}_s)$ for $s = 1, \ldots, 10,000$

- Loop over $i = 1, \ldots, 50$ times:
  - Generate a random sample of length 1,000 from the GM copula
  - Use this sample to construct a non-parametric estimate of the copula density on $(u_1 \times u_2)$ denoted $\hat{c}_i(\mathbf{u}_s)$ using one of the non-parametric models described above

- Estimate the mean integrated squared bias on $[0, 1]$ as

$$\int_{[0,1]^d} \text{Bias}(\hat{c}(\mathbf{u}))^2 d\mathbf{u} \approx \frac{1}{10,000} \sum_{s=1}^{10,000} \Big[c(\mathbf{u}_s) - E[\hat{c}_i(\mathbf{u}_s)]\Big]^2$$

- Estimate the mean integrated variance as

$$\int_{[0,1]^d} \text{Variance}(\hat{c}(\mathbf{u})) d\mathbf{u} \approx \frac{1}{10,000} \sum_{s=1}^{10,000} \text{Var}[\hat{c}_i(\mathbf{u}_s)].$$

The results for squared bias, variance and IMSE shown in Figures 4.7 – 4.9 respectively. Notice how the trade-off between bias and variance emerges for both the checkerboard and the Bernstein approximations: as we increase n we eliminate bias, at the expense of increasing variance. We can see that the optimal choice of n for the Checkerboard approximation is around 5, and around 15 for the Bernstein approximation. Notice how the Bernstein approximation has similar bias but far lower variance then the Checkerboard approximation. This is consistent with the analytic results of [San 04], who show that the variance in the Bernstein approximation grows proportional to n, while the variance in the checkerboard approximation is proportional to n^2. Note that this underlines our reason for smoothing the empirical copula, which is equivalent

to the Checkerboard approximation with $n = T$, before it can be applied as a non-parametric model of dependence. Determining the optimal value of by simulation is extremely computationally expensive, but we can see that the much more efficient LCV method specifies the optimal value of n successfully for the Bernstein estimator. We can also see that the kernel approximations compare quite favourably with the partition of unity approach: the Bernstein approximation has lower IMSE when the optimal value of n is used, but it is quicker to estimate a reasonable value for h using either (4.12) or (4.13).

In principle we can apply the non-parametric copula methods in arbitrarily many dimensions, subject only to the limits of computational power in estimating the parameters. In practice, however, we run into the 'curse of dimensionality' – a rapid decline in the quality of the non-parametric approximation as the dimension of the problem increases. This problem, which arises in all forms of high-dimensional non-parametric estimation, is discussed in detail by [Sil 86] and must be regarded as a significant limitation of the non-parametric copula approach. We remark, however, that Historical Simulation VaR (see §5.1.1) has emerged as an industry standard even for high dimensional risk management problems, despite being equivalent to using a partition of unity with $n \to \infty$ (or equivalently a kernel with $h \to 0$).

4.4.2 Simulation

Since we are using non-parametric methods to construct approximations of the empirical copula density, it seems natural to use the conditional sampling approach to simulation (this approach is frequently suggested in the literature, see [Fer 03] and [San 05]). Given the very large samples sizes required for effective Monte Carlo simulation, however, this procedure is likely to be of little use in practice, since calculating the inverse conditional density functions is computationally expensive. We introduce another, simpler, approach to non-parametric simulation, which we will show to be at least as fast as simulation from the various forms of parametric copula.

The simplest approach to non-parametric simulation is simply to draw with replacement from the empirical sample, a statistical procedure known as *bootstrapping*. This approach suffers from the problem that we will never sample from outside the original data set, which makes it difficult to, for instance, estimate the quantiles of a distribution at a high or low confidence levels. Given a multivariate kernel k with smoothing parameter h, [Sil 86] suggests the following generalisation:

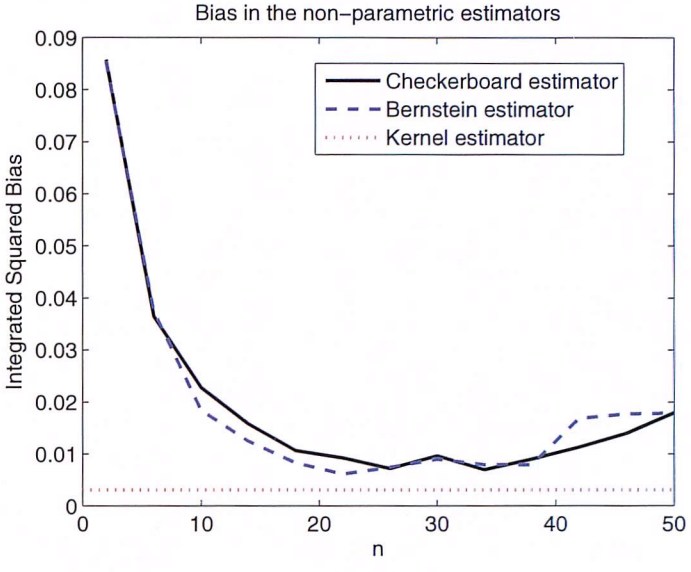


Figure 4.7: Integrated Square Bias in the non-parametric estimators

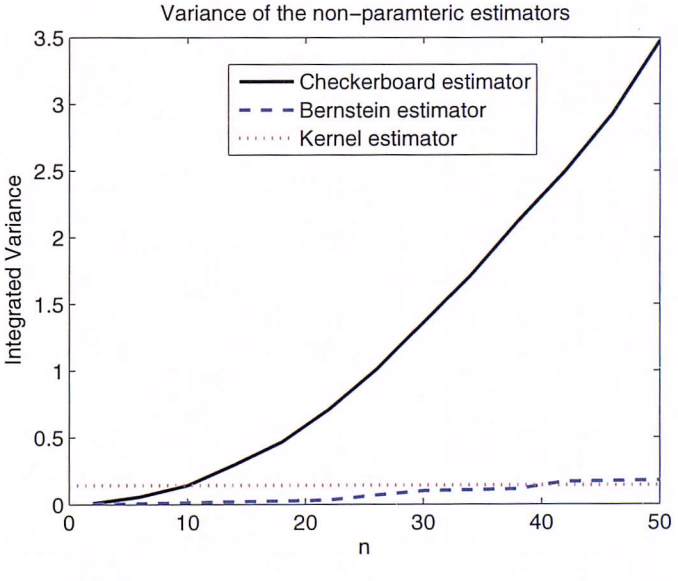


Figure 4.8: Integrated Variance in the non-parametric estimators

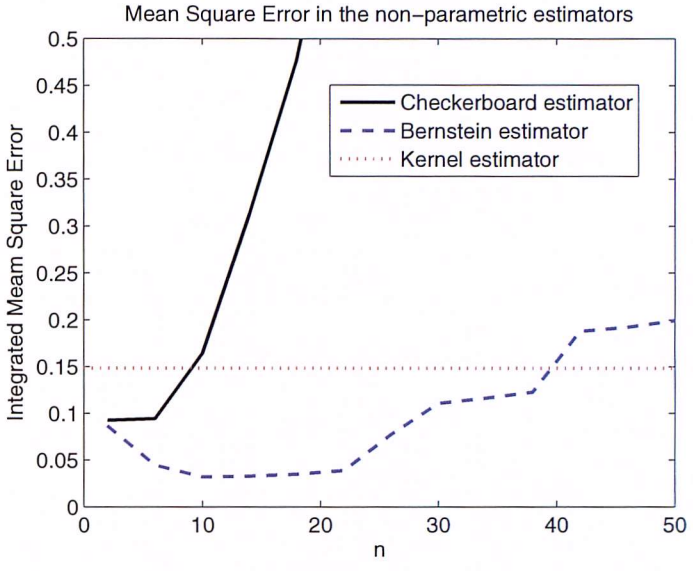


Figure 4.9: Integrated Mean Square Error in the non-parametric estimators

Kernel Simulation Algorithm

- Simulate from the uniform distribution; $v \sim U(0,1)$
- Use this to choose an integer between 1 and T; $I = \lceil vT \rceil$
- Simulate from the distribution of the multivariate kernel; $\varepsilon \sim k$
- Add the perturbation ε to the Ith element of the sample; $\mathbf{Y} = \bar{\mathbf{x}}^{(I)} + h\varepsilon$

$\mathbf{Y}$ is then a sample from the estimated density $\hat{f}(\mathbf{x}; h)$. This approach amounts to bootstrapping with a perturbation, where we use the theory of non-parametric density estimation to apply the perturbation in an appropriate fashion. Generating the sample ε from the kernel is in practice straightforward if we use a standard form of probability density function as the kernel. Using a Gaussian kernel, for example, we need only sample from a standard Gaussian distribution. As observed by [Hör 00] pg. 675, this conceptually and practically simple approach to simulation from kernel estimators "seems to be practically unknown among simulation practitioners", despite having appeared in the density estimation literature since at least [Dev 85]. This method can easily be adapted to sample from the kernel copula (4.10) or (4.11). We simply sample from the estimated multivariate kernel density using the above approach, and transform this by applying the Gaussian distribution function.

Kernel Copula Simulation Algorithm

- Simulate from the uniform distribution; $v \sim U(0,1)$
- Use this to choose an integer between 1 and T; $I = \lceil vT \rceil$
- Simulate from the distribution of the multivariate kernel; $\varepsilon \sim k$
- Add the perturbation ε to the Ith element of the sample; $\mathbf{Y} = \bar{\mathbf{x}}^{(I)} + h\varepsilon$
- Make uniform on $[0,1]^d$;

$$u_i = \frac{1}{T}\sum_{t=1}^{T} \Phi((Y_i - \bar{x}_i^{(t)})/h_i)$$

The idea of bootstrapping with a perturbation can be extended to sample from the partition of unity estimators, where the partitions of unity act like kernels with unequal weights given by $\Delta_{i_1,\ldots,i_d}$ (see (4.5) and (4.8)) and the elements of the Partition of Unity act as kernels. The problem now is to choose a particular element of the Partition of Unity by sampling from $\Delta_{i_1,\ldots,i_d}(C)$; and then to sample from the Partition of Unity itself, $\phi_{i,n}$.

For the empirical copula (in our definition), sampling from the distribution is trivial since we need only draw with replacement from the transformed empirical dataset $\{\bar{u}_1^{(t)}, \ldots, \bar{u}_d^{(t)}\}$; and in this case the Partition of Unity is simply a series of uniform distribution functions over the interval $\left[\frac{i-1}{n}, \frac{i}{n}\right]$.

Empirical Copula Simulation Algorithm

- Simulate from the uniform distribution; $v \sim U(0,1)$
- Use this to choose an integer between 1 and T; $I = \lceil vT \rceil$
- Select a uniform distribution in the first dimension based on the Ith empirical sample; $j_1 = \bar{u}_1^{(I)}(T+1)$

  $\vdots$

- Select a uniform distribution in the dth dimension based on the Ith empirical sample; $j_d = \bar{u}_d^{(I)}(T+1)$
- Simulate from the j_1th uniform distribution; $u_1 \sim \chi_{j_1,T}$

  $\vdots$

- Simulate from the j_dth uniform distribution; $u_d \sim \chi_{j_d,T}$

Samples can be drawn from the order n Checkerboard approximation and Bernstein approximation by first sampling from the empirical copula, then using this sample to 'choose' one of the n elements of the partition. Notice how in the Bernstein algorithm we used the equivalence between Bernstein polynomials and beta distributions to simplify the process, since efficient sampling techniques for the beta distribution already exist, see [Dev 86] §9.4.

Checkerboard Approximation Simulation Algorithm

- Simulate from the empirical copula; $v \sim \hat{C}$
- Use this to select a uniform distribution in the first dimension; $j_1 = \lfloor v_1(k+1) \rfloor$

$\vdots$

- Use this to select a uniform distribution in the dth dimension; $j_d = \lfloor v_d(k+1) \rfloor$
- Simulate from the j_1th uniform distribution; $u_1 \sim \chi_{j_1,n}$

$\vdots$

- Simulate from the j_dth uniform distribution; $u_d \sim \chi_{j_d,n}$

Bernstein Approximation Simulation Algorithm

- Simulate from the empirical copula; $v \sim \hat{C}$
- Use this to select a beta distribution in the first dimension; $j_1 = \lfloor v_1(k+1) \rfloor$

$\vdots$

- Use this to select a beta distribution in the dth dimension; $j_d = \lfloor v_d(k+1) \rfloor$
- Simulate from the j_1th beta distribution; $u_1 \sim \beta_{j_1+1,k-j_1+1}$

$\vdots$

- Simulate from the j_dth beta distribution; $u_d \sim \beta_{j_d+1,k-j_d+1}$

We performed a simple experiment to test the speed of the various simulation algorithms for a range of parametric and non-parametric copulas, the results are reported in Table 4.1. In each case we implemented the algorithm described in the text in the MATLAB software package using a Pentium IV 3 GHz 512 MB machine. For comparison, we have also included the time to generate a d-dimensional sample of S independent uniform random numbers (i.e. a sample from the copula $C^{\perp}$).

	$S = 1 \times 10^3$		$S = 1 \times 10^5$	
	$d = 2$	$d = 10$	$d = 2$	$d = 10$
Independent	0.0001	0.0005	0.0093	0.0499
Gaussian	0.0012	0.0048	0.1135	0.5165
Student's t	0.0099	0.0451	0.9976	4.7765
Clayton	0.0024	0.0582	0.2188	6.0409
Gumbel	0.0047	0.0142	0.4672	1.4779
Gaussian Mixture	0.0047	0.0191	0.4820	1.9992
Empirical	0.0004	0.0015	0.0407	0.1670
Checkerboard	0.0007	0.0028	0.0735	0.3109
Bernstein	0.0023	0.0101	0.2391	1.0502
Kernel	1.0173	6.9098	89.4894	392.0810

Table 4.1: Time (in s) to simulate from a range of copula models

We can see that the simulation algorithms for the Partition of Unity-based estimators are very efficient indeed: samples can be generated from the Checkerboard approximation faster than from any existing parametric copula, and the Bernstein approximation is only a little slower. The Gaussian Mixture is also very efficient, but the kernel is somewhat slower, due to the requirement to repeatedly evaluate the one-dimensional Gaussian distribution function, but even here samples can be generated quickly enough to make large scale Monte Carlo simulation feasible. Also notice that all of the simulation algorithms are at worst linear in S and d; the run-time does not 'blow up' as the number of scenarios or the dimensionality increases.

Chapter 5

Applications of Copulas in Quantitative Finance

5.1 Portfolio Risk Management

First we shall examine how copulas can be used to capture the risk exposure of some portfolio of financial instruments. We focus on market risk, though copula models can also be used to model the dependency between adverse credit events, see, for example, [Li 00] and [Sch 03b] §10.

5.1.1 Risk Measures

The essence of quantitative risk management lies in the estimation of the profit-and-loss (P&L) distribution of some portfolio over a given time horizon. Of particular interest are the tails of this distribution, where the largest gains and losses occur. Shortly we shall consider how to estimate the distribution of the portfolio profit and loss, δP, but first let us suppose that we have some model of the P&L distribution, $F_{\delta P}$, and wish to assign some single number as a measure of 'the risk'.

Perhaps the simplest statistical measure of the level of risk is the variance of δP. The intuition behind this choice of risk measure is clear (high variability implies high risk), and the consequences for portfolio selection are well known (see, for example, [Mar 52]). The principal drawback of this approach is that variance refers to the whole of the distribution, rather than the lower tail specifically (i.e. the region where large losses are made). Thus it is possible to have two portfolios with the same variance but which have a very different likelihood of exposing the holder to large losses.

An alternative risk measure which is intended to capture the likelihood of big losses

is the *Value-at-Risk*, or VaR. Defined as "the maximum possible loss over the following T days at a confidence level of $100(1-\alpha)\%$", we can write VaR mathematically in terms of the distribution of δP as

$$VaR_\alpha(\delta P) \equiv -F^{-1}_{\delta P}(\alpha).$$

The negative sign in the above definition is included to observe the convention that VaR should be a positive number. Notice that since VaR refers specifically to some quantile of the P&L distribution, usually set at 95% or greater, it captures the impact of large losses, as we might wish. Value-at-Risk has become the most widely used measure of portfolio risk by major financial institutions, not least because it has been codified in regulatory capital rules (see [Bas 05]) and because a number of 'backtesting' techniques have been developed to test the accuracy of a VaR estimate, see §5.1.3.

VaR has received some criticism in recent years for its failure to satisfy the property of subadditivity [Art 99]. This means that for certain portfolios VaR fails to respect diversification: for two portfolios A and B, we expect that the risk on the combined portfolio should be smaller than the risk on each portfolio added together, i.e. $\text{risk}(A+B) \leq \text{risk}(A) + \text{risk}(B)$. It is possible to give examples of portfolios such that this fails to hold when VaR is used a risk measure, leading [Art 99] to propose a new risk measure known as *Expected Shortfall*, or ES:

$$ES_\alpha(\delta P) \equiv -E[\delta P | \delta P < VaR_\alpha].$$

It can be shown that this measure satisfies subadditivity in all cases, as well as number of other properties we might consider desirable. ES has not, however, been widely adopted by the industry as a measure of market risk. This is probably because examples of portfolios with non-subadditive market risk are in practice rare, and because it is not clear how to perform a statistical check on an estimate of ES. We shall use VaR as a measure of risk in what follows.

5.1.2 Modelling Market Risk

As mentioned above, estimating portfolio risk requires us to estimate the distribution of returns, δP. For this we require a two things. First, a statistical model for the behaviour of the d market factors to which we are exposed, which we denote δY_i for $i = 1, \ldots, d$. By 'market factors' we mean data on which the value of financial assets depend, for example interest rates, equity prices and currency exchange rates. Second, we need a model for how the value of each instrument in our portfolio changes depending on the

market factors, i.e. we want to estimate $\delta P_j = \delta P_j(\delta Y_1, \ldots, \delta Y_d)$. By 'instruments' we mean actual financial assets such as bonds, stocks and derivatives. The change in value of the portfolio can be computed by summing the change in value of each of its N components, i.e.

$$\delta P = \sum_{j=1}^{N} \delta P_j$$

We will mostly be concerned with modelling δY_i, since estimating δP_j is fairly straightforward for all but the most complex of assets. Copula functions can be used to split the problem of modelling δY_i into two parts: construct a model for the marginal distributions; and a model for the dependency. A wide range of approaches to modelling the marginals have been proposed in the literature, mostly involving the fitting of some 'heavy tailed' probability distribution to empirical data (see, for example, [Rac 03]).

We have chosen to adopt a slightly different approach, and use the semi-parametric Extreme Value Theory (EVT) method proposed by [Dan 00]. Their methodology basically involves using the empirical distribution function to model the body of the distribution, and applying EVT to smooth the tails. The application of EVT is based on the result that, for the kind of 'heavy tailed' distributions normally observed in time series of financial data, the distribution function has Pareto like tails, i.e.

$$F(x) \approx a(-x)^{-\alpha_l} \quad \text{as} \quad x \to -\infty$$

and

$$F(x) \approx 1 - ax^{-\alpha_u} \quad \text{as} \quad x \to \infty$$

where $a > 0$ is a scaling constant and the tail indexes $\alpha_l > 0$ and $\alpha_u > 0$ determine the heaviness of the lower and upper tail. It can be shown (see [Hil 75]) that the tail index can be estimated from an ordered empirical sample by

$$\frac{1}{\widehat{\alpha_u}} = \frac{1}{M_u} \sum_{m=M_u}^{T} \left(\log \frac{\bar{X}^{(m)}}{\bar{X}^{(M_u+1)}} \right),$$

where M_u is some threshold value such that all $\bar{X} > \bar{X}^{(M_u)}$ are considered to be 'in the upper tail' of the distribution. An analogous estimator can be used to compute α_l, the lower tail index of the distribution. The choice of optimal thresholds M_u and M_l is the subject of some debate in the literature, we have used the slightly ad hoc but apparently effective approach of specifying the threshold as the second percentile for all six datasets. Figure 5.1 shows the tail index as function of threshold percentile for the lower and upper tail of the SG index data, here we can see that a values of $\alpha_l = 2.94$

and $\alpha_u = 3.48$ result from a choice of $M_l = 0.02 \times T$ and $M_u = 0.98 \times T$. Estimates of α_l and α_u for all of the sample data are given in Table B.1 in the appendix.

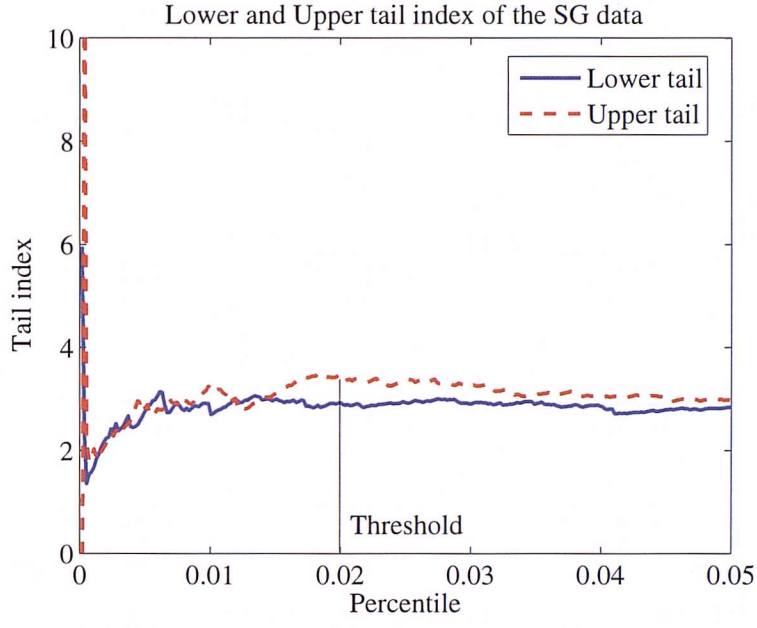


Figure 5.1: Lower and Upper tail index of SG data

Given estimates of α_u and M_u, [Dan 00] show that we can compute the pth quantile of the distribution as

$$F^{-1}(p) \equiv x_p = \bar{x}^{(M+1)} \left(\frac{Tp}{M_u} \right)^{-\alpha_u}, \tag{5.1}$$

and equivalently for the lower tail. This leads us to the following algorithm for generating Monte Carlo samples x from an ordered empirical dataset $\bar{x}^{(1)} < \bar{x}^{(2)} < \cdots < \bar{x}^{(T)}$:

EVT Sampling Algorithm

- Generate a uniform random number $u = U(0,1)$
- If u does not lie in the upper or lower tail of the distribution, i.e. $\frac{M_l}{T} < u < \frac{M_u}{T}$, then set x by linearly interpolating the empirical data
- If u lies in the upper tail, i.e. $u > \frac{M_u}{T}$, then use equation 5.1 to generate x, i.e.

$$x = \bar{x}^{(M_u+1)} \left(\frac{Tu}{M_u}\right)^{-\alpha_u}$$

- If u lies in the lower tail, i.e. $u < \frac{M_l}{T}$, then

$$x = \bar{x}^{(M_l-1)} \left(\frac{Tu}{M_l}\right)^{-\alpha_l}$$

The problem now is to join the EVT distributions in an appropriate fashion. To do this we need to draw dependent uniform random numbers $u_1, \ldots, u_d$, i.e. we need to sample from a copula. In the following section we shall use the EVT model for the marginals together with a range of copula models to estimate the risk on a portfolio of equity indexes.

5.1.3 Portfolio Experiments

We would like to examine the usefulness of the Gaussian Mixture copula as a model for the dependence between market factors, with the aim computing the market risk on some portfolio. We wish to implement the models in a way that would be realistic for a bank or other financial institution. To this end, we use a rolling calibration period of four years, and update the model parameters on a bi-monthly basis. For the marginal distributions we use semi-parametric EVT as described above. The models of dependence we test are the Gaussian copula, Student's t copula, Clayton copula, and the Gaussian Mixture copula, each calibrated as described in the relevant section of chapter 3. We have chosen not to use the non-parametric models here, since the relatively large number of underlyings means they would begin to suffer from the 'curse of dimensionality'.

The models are assessed by backtesting, a procedure for checking the accuracy of VaR estimates. The details are by described, for example, by [Dow 02] §9, but the basic idea is as follows: we measure portfolio VaR at confidence level $\alpha\,\%$ over some period of T days, and then compare the estimated VaR with the realized P&L distribution.

If the VaR model is accurate, then in a period of T days we ought to observe around $(100\% - \alpha) \times T$ VaR breaks, that is days when the actual P&L is less then the VaR. If we observe fewer than $(100\% - \alpha) \times T$ breaks it is likely that the VaR is overestimated, while if we observe more than $(100\% - \alpha) \times T$ breaks we may conclude that the VaR is too low.

The results of this exercise can be displayed visually using a backtesting chart, see Figures 5.2 - 5.4 which show time series of the VaR estimated using the various models and the actual portfolio P&L. The colored lines indicate the VaR estimated by a particular model (recall that exactly the same EVT model is used for the marginals in each case, so the differences in VaR are due purely to our choice of copula), while the black bars show the daily return on a portfolio consisting of equal-sized long positions in each of the six Asian equity indexes we have studied. Market shocks such as the crash of October 1987 and the Asian crisis of 1997-1998 can be seen clearly. There are no VaR results computed for the initial four year calibration period January 1985 - January 1989. Notice that we have also computed VaR in the positive tail of the P&L distribution. This may seem surprising (since we aim to compute the likelihood of large losses, not large gains!), but consider that the upper tail of our long portfolio will correspond exactly to the lower tail of (i.e. losses in) a short portfolio. By including upper as well as lower tails we effectively double the volume of testable data for little cost, and this is standard industry practice, see [Dow 02]. We computed P&L and VaR on every business day from 08-Jan-89 to 01-Jun-06, on days when a particular market was closed we simply assigned zero P&L and zero VaR to that component of the portfolio.

A range of confidence levels can be used to check the accuracy of the VaR at various points in the tail of the distribution. We tested the model VaR at confidence levels of 95%, 99% and 99.9%, and over a total period of 4497 days ($21\frac{1}{2}$ years less the initial 4 year calibration period). Thus we expect to observe 225, 45 and 4 VaR breaks at each confidence level respectively. Moreover, the binomial distribution can be used to place p-values on the null hypothesis that "the correct model has been used", given the number of VaR breaks actually observed. The minimum and maximum number of breaks not to reject the null hypothesis at a p-value of 5% are shown in the last row of Table 5.1.

We can see from Table 5.1 that the Gaussian copula models the portfolio loss distribution quite well at low confidence levels ($\approx 5\%$), but, as expected, seriously underestimates the risk of a large loss. The Student's t copula performs a little better, but still generally underestimates the risk, particularly in the lower tail of the loss distribution.

Model of Dependency	Lower Tail VaR			Upper Tail VaR		
	99.9%	99%	95%	5%	1%	0.1%
Gaussian	15	79	245	210	52	15
Student's t	7	71	257	216	50	10
Clayton	9	62	222	275	91	20
Gaussian Mixture	4	44	247	206	38	7
Expected # of breaks	4	45	225	225	45	4
Rejection bounds	[1 8]	[34 56]	[201 249]	[201 249]	[35 56]	[1 8]

Table 5.1: Number of VaR breaks using a range of copula models and confidence intervals

The Clayton copula performs very poorly in the upper tail of the loss distribution, which might have been expected given the lack of upper quantile dependence in this model. More surprisingly, it also underestimates the dependence in the lower tail as well. This may be due to the way we have calibrated to the median value of the Kendall's tau matrix rather than to the lower quantiles, but this underlines the key weakness of Archimedean copulas: their severe lack of free parameters with which to calibrate. The best results are achieved with the Gaussian Mixture copula: in no case does this model lead to a p-value of less than 5%, and the performance is particularly good at high confidence levels where the other models fail. Notice that the 'tail independent' Gaussian Mixture copula resulted in fewer breaks at high confidence levels than the 'tail dependent' Student's t and Clayton models - again we suggest that it is the finite sample properties of the copula that matter most, not the asymptotic behaviour of the tail of the copula.

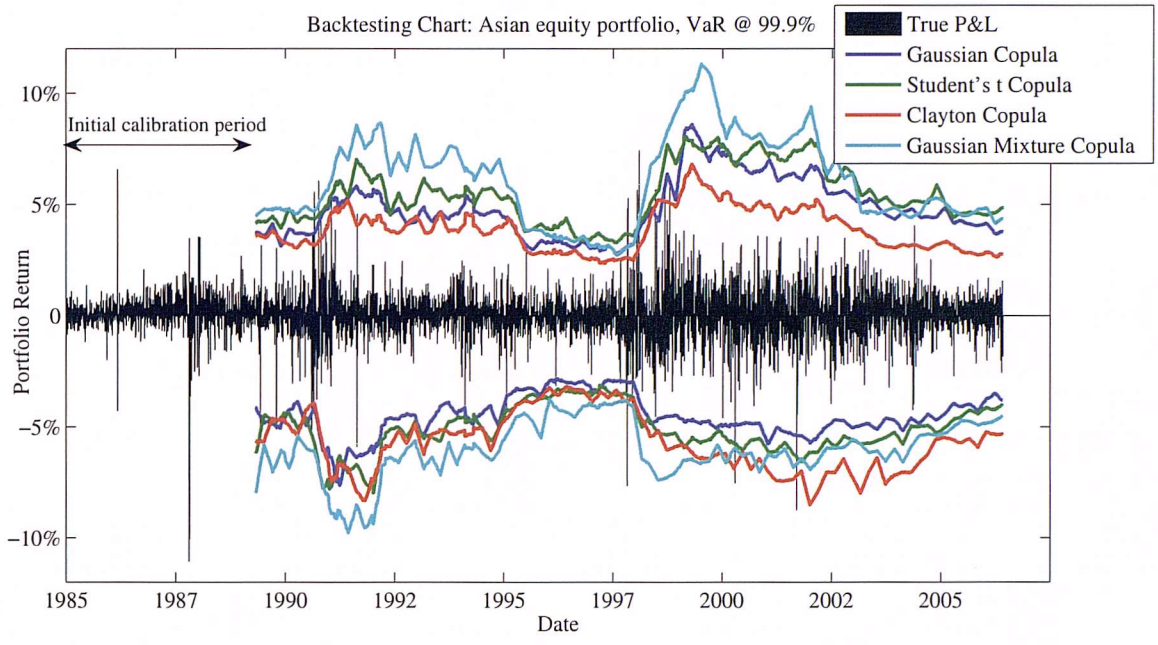


Figure 5.2: Backtesting at a confidence level of 99.9%

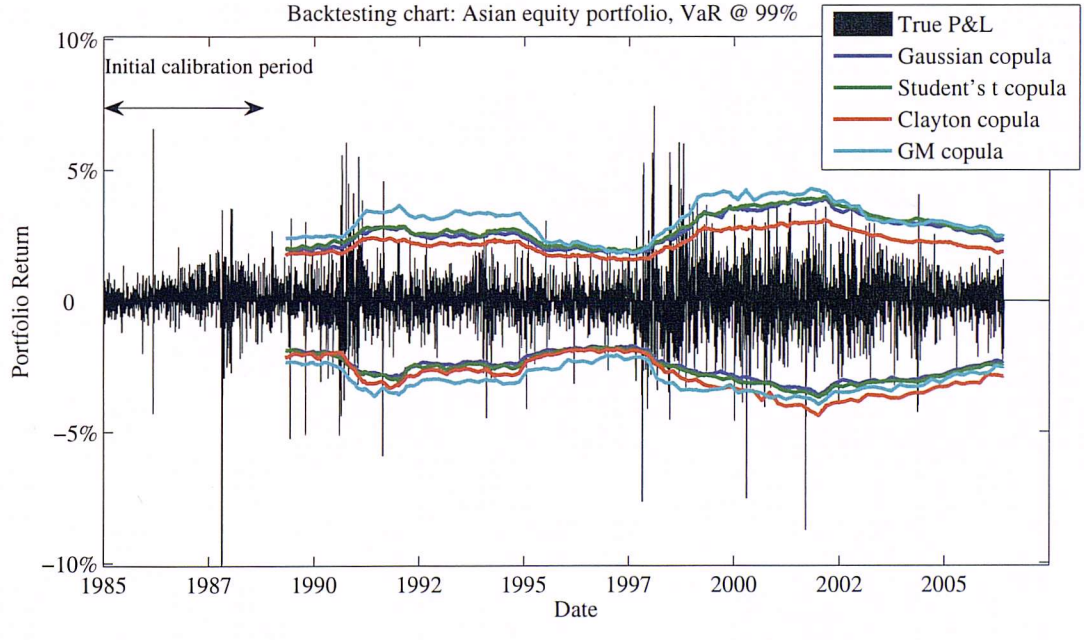


Figure 5.3: Backtesting at a confidence level of 99%

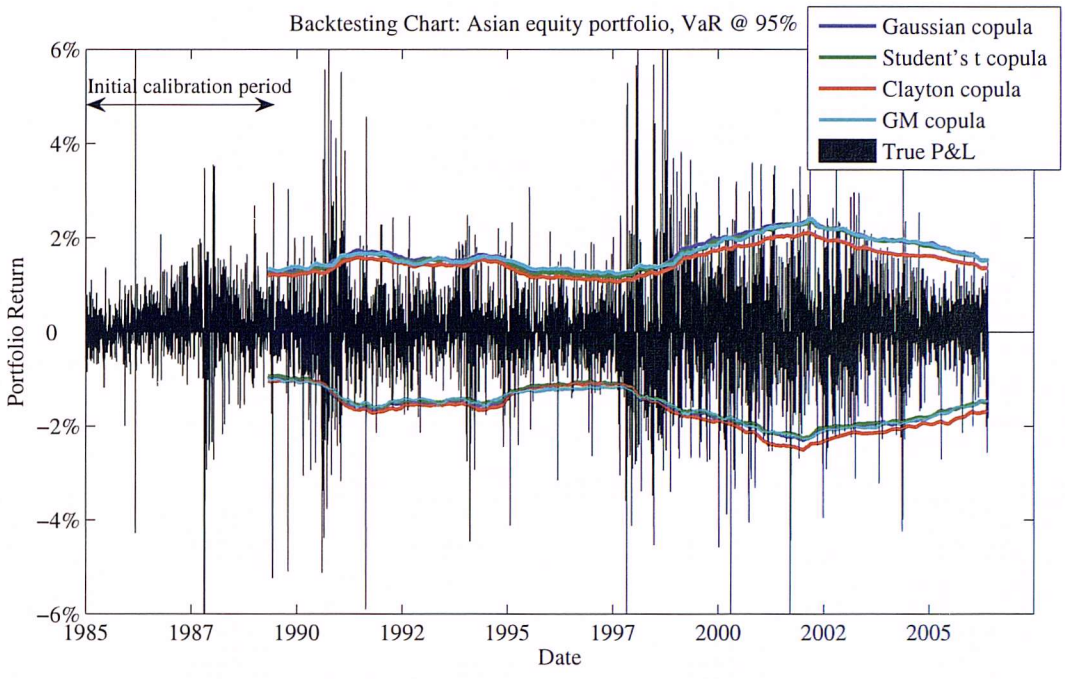


Figure 5.4: Backtesting at a confidence level of 95%

5.2 Pricing Multi-asset Derivatives

5.2.1 Pricing Theory

We assume that the reader is familiar with classical option pricing theory, and shall quote directly some standard results for the pricing of multivariate contingent claims in a complete market. For details of the theory behind these results see, for example, [Mus 05].

Consider a European contingent claim with pay-off function $G(X_1, \ldots, X_d)$ dependent on the values of the d underlyings $X_1, \ldots, X_d$ at expiry. The option pays off at time T and the deterministic risk-free interest rate is r. Using standard option pricing theory, we have that the price P of the claim at time t is the discounted expectation of the pay-off under the risk neutral measure, i.e.

$$P(X_1, \ldots, X_d, t) = e^{-r(T-t)}\mathbf{E}^*[G(X_1, \ldots, X_d)|\mathcal{F}_t]$$

$$= e^{-r(T-t)} \int \cdots \int G(X_1, \ldots, X_d) f^*(X_1, \ldots, X_d) dX_1 \ldots dX_d.$$

Here $*$ denotes the martingale probability measure, and hence $f^*(X_1, \ldots, X_d)$ denotes the multivariate density function of the underlyings in a risk-neutral setting.

We may use (2.4) to write the above expression in terms of copulas as

$$f^*(X_1, \ldots, X_d) = f_1^*(X_1) \cdots f_d^*(X_d) c^*(F_1^*(X_1), \ldots, F_d^*(X_d))$$

where c^* denotes the risk-neutral copula density. [Ros 03] shows that, under some fairly general conditions on the pricing kernel, the copula in the risk neutral measure coincides with the copula in the statistical measure, i.e.

$$c^*(F_1^*(X_1), \ldots, F_d^*(X_d)) = c(F_1(X_1), \ldots, F_d(X_d)).$$

This assumption has also been used by [Goo 05], and implies that while the risks of the marginals are priced by the market, the price of risk for a given outcome of X_i is not affected by the outcome of X_j. [Luc 05] dismiss this approach as "adhockery", but their proposed solution is to calibrate a multivariate model for the underlyings solely to univariate option prices, and implicitly assume some dependency structure. Since we can only extract information on the dependency structure from multivariate historical data or multivariate option prices, this approach seems unlikely to be fruitful. We shall stick to the assumption that the statistical and risk neutral copulas coincide. Hence we can now express the pricing formula as

$$\begin{aligned} P(X_1, \ldots, X_d, t) = e^{-r(T-t)} \int \cdots \int G(X_1, \ldots, X_d) \\ f_1^*(X_1) \cdots f_d^*(X_d) c(F_1(X_1), \ldots, F_d(X_d)) dX_1 \ldots dX_d. \end{aligned} \tag{5.2}$$

The pricing of multi-underlying options can thus be broken into three steps:

1. Construct some risk neutral model for the marginal distribution of each underlying
2. Find a model for the copula function linking the marginals in the statistical measure
3. Evaluate the d-dimensional integral in (5.2).

The calibration of the risk neutral marginals can be done using the traded prices of univariate options on each underlying. There are many ways of doing this by fitting some parametric (e.g. [Rit 90]) or non-parametric model (e.g. [Aït 98]). We adopt a relatively simple approach and assume that the marginals follow a log-normal distribution, i.e. we use the Black-Scholes Model [Bla 73] for the evolution of asset prices.

Definition 25 *Under the* Black-Scholes Model, *the* Market *is comprised of d risky assets,* $X_i(t)$, $i = 1, \cdots, d$, *which evolve according to a Geometric Brownian Motion:*

$$dX_i(t) = \mu_i X_i(t) dt + \sigma_i X_i(t) dW_i(t),$$

where $dW_i(t)$ denotes a Wiener process. The constant parameters μ_i and $\sigma_i > 0$ respectively denote the drift and the volatility of the ith asset, and the Wiener processes have pairwise linear correlations

$$\rho_{ij} = E[dW_i(t)dW_j(t)].$$

We are well aware of the limitations of this as an empirical model of the capital markets (see, for example, [Bla 88]), but we have chosen to use it for two main reasons. First, it is still the most commonly used option pricing model, at least for multi-asset options, and all of the analytic pricing formulae available for such options are derived using this approach (see the next section). Second, our main goal here is to analyse the impact of the dependency structure on option prices, rather than to construct the best possible pricing model in every respect.

For the estimation of the objective copula we shall calibrate to time series of the underlying returns at the same time horizon as the option tenor. Obviously this limits the size of the calibration set when the time to expiry is long, for now we look at pricing options close to expiry. We shall construct option prices using a variety of parametric and nonparametric copula models and compare them with the prices in the Black-Scholes (i.e. Gaussian copula) model. Where closed-form solutions are not available (i.e. whenever we do not apply the Gaussian copula model) we shall use Monte Carlo simulation to evaluate the d-dimensional integral in (5.2).

5.2.2 Option Pricing Formulae

We shall examine the three principal classes of multi-asset option: Basket options, Spread options and Best-of/Worst-of options – we place options on the Best-of and Worst-of in the same class since their prices are closely linked by no-arbitrage relationships. For each type of option there exist (sometimes approximate) closed form solutions for the price using the Black-Scholes model. First of all let us recall the pricing formula for options with a single underlying.

Call & Put Options

Perhaps the simplest form of traded contingent claims are European Call and Put options. These have pay-off at expiry of

$$G_{\text{Call}} = \max(X(T) - K, 0) \qquad \text{and} \qquad G_{\text{Put}} = \max(K - X(T), 0)$$

where K is referred to as the strike price of the option.

In the Black-Scholes model, one can derive the famous Black-Scholes [Bla 73] formula for the prices of univariate European Call and Put options written on the underlying X

$$P_{\text{Call}} = X(t)\Phi\left[d_1(X(t), K, \sigma)\right] - Ke^{-r(T-t)}\Phi\left[d_2(X(t), K, \sigma)\right] \tag{5.3}$$

and

$$P_{\text{Put}} = Ke^{-r(T-t)}\Phi\left[-d_2(X(t), K, \sigma)\right] - X(t)\Phi\left[-d_1(X(t), K, \sigma)\right], \tag{5.4}$$

where

$$d_1(X(t), K, \sigma) = \frac{\log(X(t)/K) + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}$$

and

$$d_2(X(t), K, \sigma) = \frac{\log(X(t)/K) + \left(r - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}.$$

Since K and T are specified in the contract and $X(t)$ and r may be observed directly in the market, the only 'free parameter' in the Black-Scholes formula is the volatility, σ. The value of an option increases with increased volatility, since this will increase the likelihood of the price of the underlying ending 'in the money'.

We now define three classes of rainbow option, and try to form some intuition about how the price of each is influenced by the dependency structure. This is examined by [Fen 03] in a similarly intuitive way, and more formally by [Rap 01]. Note that both of these authors limit their attention to the Gaussian copula, whereas we shall investigate the impact of a range of copula models in §5.4. We also mention the work of [Ros 03], who used a non-parametric (kernel) copula to find the prices of univariate options on the GBP-JPY rate that are implied by the joint distribution of the GBP-USD and JPY-USD exchanges rates. The 'triangular' relationship exhibited by currency pairs mean that in this particular case option prices can be computed without Monte Carlo simulation, for which efficient non-parametric techniques were not previously available.

Basket Option

A Basket option has pay-off dependent on the sum (or equivalently the average) of the price of its constituents:

$$G_{\text{Call-on-Basket}} = \max(X_{\text{B}}(T) - K, 0),$$

$$G_{\text{Put-on-Basket}} = \max(K - X_{\text{B}}(T), 0),$$

where $X_{\text{B}}(t) = X_1(t) + \cdots + X_d(t)$.

Using the a Black-Scholes model one can obtain a closed form price for basket options by assuming that the value of the basket is log-normally distributed at expiry, i.e. assuming that the sum of log-normal variables is itself log-normally distributed. While this is not strictly accurate, it has been shown to give a good approximation to the price of options dependent on the arithmetic average of log-normal variables (see [Hul 03], §19.12). Using the properties of the log-normal distribution, it can be shown that the first two moments of X_{B} in a risk-neutral world at time T are

$$M_1(X_B,T) = e^{r(T-t)}\sum_{i=1}^{d} X_i(t) \quad \text{and} \quad M_2(X_B,T) = e^{2r(T-t)}\sum_{i=1}^{d}\sum_{j=1}^{d} X_i(t)X_j(t)e^{\sigma_i\sigma_j\rho_{ij}},$$

and that the volatility of the basket is then

$$\sigma_{\mathrm{B}} = \sqrt{\frac{1}{T-t}\ln\left(\frac{M_2}{M_1^2}\right)}.$$

Assuming that the basket's value at expiry is log-normally distributed, we can simply set $X(t) = X_{\mathrm{B}}(t)$ and $\sigma = \sigma_{\mathrm{B}}$ in (5.3) and (5.4) to price Basket-Call and Basket-Put options.

The price of a basket option can be expected to increase with increasing dependency, since the volatility of the basket will increase as its constituents tend to move up and down together. [Qu 06] has studied the pricing of basket options outside of the Gaussian framework, using a non-parametric distribution for the marginals but retaining a Gaussian copula for the dependency. He reports that the volatility skew implied by the non-parametric models for the marginals is carried over to the price of the basket, but that the basket skew is less pronounced than that of the underlyings. He also shows that the basket skew is most pronounced when the dependency between the marginals is strongest.

Spread Option

A spread option pays off on the difference in price between a pair of underlyings:

$$G_{\text{Call-on-Spread}} = \max(X_1 - X_2 - K, 0),$$

$$G_{\text{Put-on-Spread}} = \max(K - X_1 + X_2, 0).$$

The problem of pricing this type of option in a Black-Scholes setting was first considered by [Mar 78] who developed a closed form solution for an option to exchange one asset for another, which is equivalent to a zero-strike spread option. The following closed

form approximations to the prices of Call-on-Spread options with $K \neq 0$ were given by [Kir 95].

$$P_{\text{Call-on-Spread}} = X_1(t)\Phi\left(\frac{\log(X_S)}{\sigma_S} + \frac{\sigma_S}{2}\right) - (X_2(t) + Ke^{-rT})\Phi\left(\frac{\log(X_S)}{\sigma_S} - \frac{\sigma_S}{2}\right),$$

where

$$X_S(t) = \frac{X_1(t)}{X_2(t) + Ke^{-rT}}$$

and

$$\sigma_S = \sqrt{\sigma_1^2 + \left(\sigma_2 \frac{X_2(t)}{X_2(t) + Ke^{-rt}}\right)^2 - 2\rho\sigma_1\sigma_2 \frac{X_2(t)}{X_2(t) + Ke^{-rt}}}.$$

The price of a spread option can be expected to fall with increasing dependency, as the likelihood of the underlyings moving apart is reduced. The problem of using copulas to price zero-strike spread options was considered by [Che 02], who used a Frank copula to show that the price of such options deceases with increased dependency.

Best-of/Worst-of Options

Finally we consider options on the maximum and the minimum of a selection of underlyings

$$G_{\text{Call-on-Best}} = \max(\max(X_1, \ldots, X_d) - K, 0)$$

$$G_{\text{Put-on-Best}} = \max(K - \max(X_1, \ldots, X_d), 0)$$

$$G_{\text{Call-on-Worst}} = \max(\min(X_1, \ldots, X_d) - K, 0)$$

$$G_{\text{Put-on-Worst}} = \max(K - \min(X_1, \ldots, X_d), 0).$$

In pricing these options it is useful to note the following no-arbitrage relationships, all of which follow from the equality of pay-off at expiry,

$$\begin{aligned}
P_{\text{Call-on-Worst}}(X_1, X_2; K) &= P_{\text{Call}}(X_1; K) + P_{\text{Call}}(X_2; K) - P_{\text{Call-on-Best}}(X_1, X_2; K), \\
P_{\text{Put-on-Best}}(X_1, X_2; K) &= P_{\text{Call-on-Best}}(X_1, X_2; K) - P_{\text{Call-on-Best}}(X_1, X_2; 0) + Ke^{-r(T-t)}, \\
P_{\text{Put-on-Worst}}(X_1, X_2; K) &= P_{\text{Call-on-Worst}}(X_1, X_2; K) - P_{\text{Call-on-Worst}}(X_1, X_2; 0) + Ke^{-r(T-t)}.
\end{aligned}$$

We have limited the above expressions to two dimensions for clarity, but the d-dimensional extensions follow trivially. The problem of pricing Call or Put options on the maximum or minimum of a basket is hence reduced to pricing a Call-on-Best option plus univariate options on each underlying. This problem was solved for the Black-Scholes model by [Stu 82] for the bivariate case and [Joh 87] for the multivariate case. For conciseness we only quote the bivariate Call-on-Best formula

$$P_{\text{Call-on-Best}} = X_1(t)\Phi_2\left[d_1(X_1(t), K, \sigma_1), d_1'(X_1(t), X_2(t), \sigma_{12}); \rho_{112}\right]$$

$$+X_2(t)\Phi_2\left[d_1(X_2(t),K,\sigma_2),d_1'(X_2(t),X_1(t),\sigma_{12});\rho_{212}\right]$$
$$-Ke^{-r(T-t)}\left[1-\Phi_2(-d_2(X_1(t),K,\sigma_1),-d_2(X_2(t),K,\sigma_2);\rho_{12})\right].$$

where

$$d_1'(X_i,X_j,\sigma)=\frac{\log(X_i/X_j)-\frac{1}{2}\sigma^2(T-t)}{\sigma_i\sqrt{T-t}},\qquad \sigma_{ij}=\sqrt{\sigma_i^2+\sigma_j^2-2\rho_{ij}\sigma_i\sigma_j}$$

and

$$\rho_{ijk}=\frac{\sigma_i^2-\rho_{ij}\sigma_i\sigma_j-\rho_{ik}\sigma_i\sigma_k-\rho_{jk}\sigma_j\sigma_k}{\sigma_{ij}\sigma_{ik}}.$$

It is interesting to note that these expressions involve the multivariate normal distribution function, in contrast to the solutions for spread and basket options which can be expressed as single integrals.

The price of a Call-on-Best option is, *ceteris paribus*, decreasing in dependency. We can intuitively understand this as being caused by dispersion in the distribution - as the underlyings become less correlated, the chance a particular underlying taking a high value increases, and hence the typical value of $\max(X_1,...,X_d)$ increases. Thanks to the no-arbitrage relations above, we can deduce the effect of increasing dependency on the other types of Best-of and Worst-of option: increased dependency increases the price of Put-on-Best and Call-on-Worst options, while decreasing the price of a Put-on-Worst.

A couple of authors ([Che 02] and [Goo 05]) have used copulas to price basket options. The former used a Frank copula to price bivariate baskets, but do not show how their results compare with those from any other model. The latter used a selection of Archimedean models, which they found to give generally higher prices for bivariate Call-on-Max and Put-on-Min options than a Gaussian model with the same value of Kendall's tau. They report "less significant" price differences for Call-on-Min and Put-on-Max options, which is difficult to understand given the no-arbitrage relationships that exist between the four classes of basket option.

5.2.3 Pricing with Copulas

We would like to examine the effect that out choice of copula has on the price of multivariate options with a range of strikes. We will use consider the following examples: a Call-on-Best and Call-on-Worst option on the Nikkei 225 and Taiwanese indexes; a Call-on-Spread option on the Hang Seng and Singapore indexes; and a Call-on-Basket option on the Nikkei 225, Kospi and Bangkok Book Club indexes. In each case we normalize the underlyings to have price at $t=0$ of $X_i(0)=\$100$. Normalization so that the starting values are approximately equal is required to make these products meaningful (a best-of

option written on 01-Jun-2006 on the absolute value of the Nikkei 225, spot price = 15503, and the Kospi index, spot price = 1295, would attract few buyers, for example), and is standard practice in the market. In each case we consider options with one month to expiry, and calibrate to non-overlapping one month returns from the data, giving a calibration set with between 250 and 300 data points. The marginal distributions are assumed to be log-normally distributed with constant annualized volatility $\sigma = 0.25$, and we assume that the annualized risk free rate of interest is $r = 5\%$. We examine the behaviour of each product over a range of strikes: $K = \$270$ to $\$340$ for the Call-on-Basket; $K = -\$15$ to $\$15$ for the Call-on-Spread; and $K = \$80$ to $\$120$ for the Call-on-Best and Put-on-Worst.

We compare four models for the dependency structure:

- Gaussian copula (i.e. the standard Black-Scholes model) with correlation calibrated using equation (3.1)
- Student's t copula, calibrated as described in §3.1.2.
- Clayton copula, calibrated using the Kendall's Tau equation from Table 3.1
- Bernstein approximation with n chosen by LCV, as described in §4.4.1

We have chosen to use the Bernstein approximation as a non-parametric model for dependency since it it exhibits much lower mean-squared error the Checkerboard approximation (see §4.4.1), and is more efficient for large scale simulations than the kernel (see §4.4.2).

Where a Gaussian dependency structure is used, options are priced used the formulas in the previous section, in all other cases options are priced using Monte Carlo simulation with 1,000,000 simulations. We report our results both in terms of option price and in terms of the 'implied correlation', i.e. the value of the correlation parameter which *ceteris paribus* can be used in a log-normal model to recover the price implied by a given copula.

Ideally we would like to compare the prices and implied correlations from each copula with the prices of traded options, but unfortunately this has not been possible since there is no liquid market for multi-underlying options. Our pricing results must therefore be viewed as an experiment in the sensitivity of prices in the Black-Scholes model to a more 'realistic' dependence structure. For this reason we view the results generated by the non-parametric copula models to be of particular interest, since they may reasonably said to reflect the 'true' joint distribution of returns.

Basket Options

Results for a call option on basket comprised of the Nikkei 225, Kospi and Bangkok Book Club indexes are shown as a function of strike in Figures 5.5 and 5.6. The first shows the difference in between the Black-Scholes price of the option (with ρ calibrated to the historical value of τ) and the price implied by each of the copula models. The second figure shows the implied correlation from each model, i.e. the single value of ρ under which the Black-Scholes price is equal to that implied the chosen copula.

The Student's t copula increases the price of in-the-money and out-of-the-money options, while reducing the price of at the money products. We can understand this by considering the nature of the t-copula: increased dependence in the upper and lower tails means the likelihood of a large move in the basket price is increased, increasing the value of out-of-the-money and in-the-money options. This is similar to the effect commonly seen in the market for vanilla equity options, where prices at strikes much lower and much greater than the spot are frequently higher than the Black Scholes formula would suggest based on observed volatility. In this sense, the implied correlation shown in Figure 5.6 for the t copula is analogous to the volatility smile seen in the market for vanilla options.

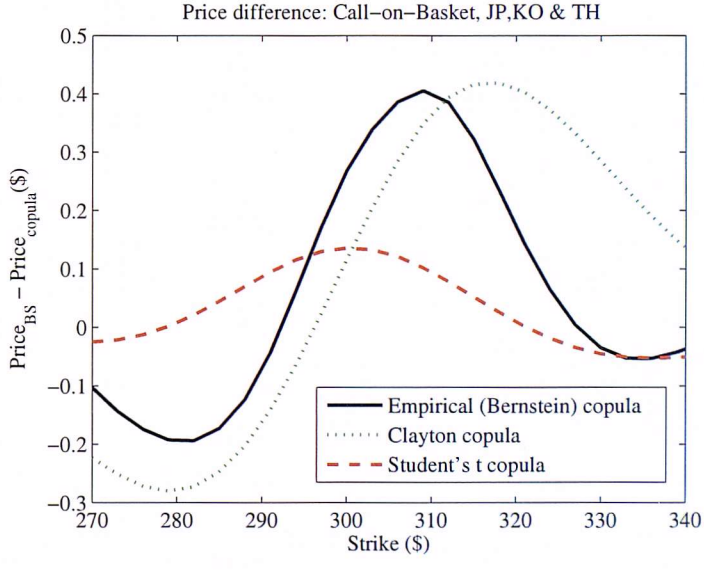


Figure 5.5: Absolute price differences for a Call-on-Basket option

The Clayton copula, in contrast, generates a 'correlation skew', i.e. out-of-the-money options are more expensive than in the Black Scholes model, but in-the-money baskets

are cheaper. Again this can easily be explained by the nature of the Clayton model: large downward moves in the price of the basket are made more likely, while large upward movements are very unlikely to be observed, due to the low upper quantile dependence of this copula. Also notice that the Clayton copula has a greater impact on prices than the Student's t, which is perhaps unsurprising since the Student's t is 'closer' to the Gaussian model than the Clayton.

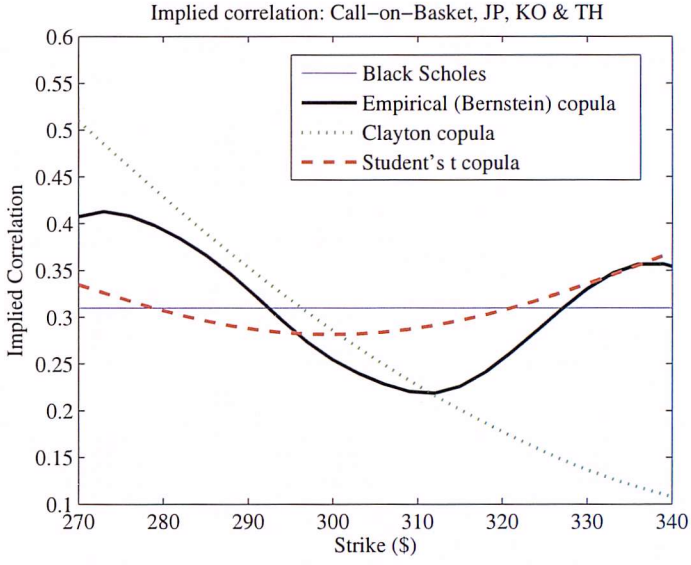


Figure 5.6: Implied correlations for a Call-on-Basket option

The results for the non-parametric model can be viewed as a mixture of the effects seen for the Student's t and the Clayton. For at-the-money and out-of-the-money options the behaviour is similar to (if a little less pronounced than) that of the Clayton: thanks to lower quantile dependence in the model, at-the-money options are cheaper and out-of-the-money options are more expensive than in the Black-Scholes model. For in-the-money options, the results are closer to those for the Student's t model: a reflection of the mild upper quantile dependence in the data.

Spread Options

Pricing results for a Call-on-Spread option on the HK and SG indexes are shown in Figures 5.7 and 5.8 In this case the results for the Clayton and the Student's t copulas are similar: they both imply that spread options with non-zero strike are under-priced by the Black-Scholes model. This is due to an interesting, but little studied property of these models: not only do they exhibit greater lower and(in the Student's t case) upper

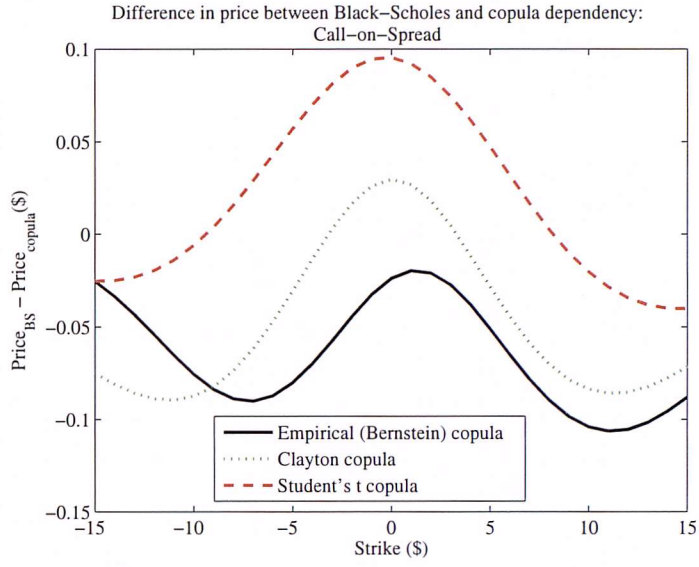


Figure 5.7: Absolute price differences for a Call-on-Spread option

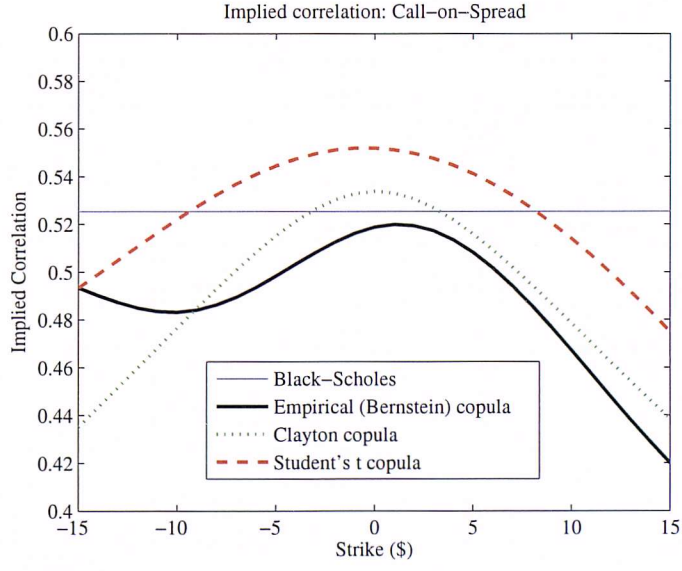


Figure 5.8: Implied correlations for a Call-on-Spread option

quantile dependence than the Gaussian copula, they imply that large opposite moves occur more frequently than the Gaussian model implies, i.e. for small values of u,

$$c_{\mathrm{t}}(u, 1-u) > c_{\mathrm{Ga}}(u, 1-u) \quad \text{and} \quad c_{\mathrm{Cl}}(u, 1-u) > c_{\mathrm{Ga}}(u, 1-u).$$

This has the effect of inducing fat tails in the distribution of spreads, which raises prices of away-from-the-money options. This creates a 'correlation frown', since the price of spread options is a decreasing function of dependency. A similar effect is observed for the empirical approximation, suggesting that the Gaussian model underestimates the likelihood of a large opposite move as well as a large joint move in these indexes.

Best-of and Worst-of Options

Prices[1] and implied correlations for a Call-on-Best and a Call-on-Worst option on the JP and TW indexes are shown in Figures 5.9 to 5.12. For the Call-on-Best option, the low upper quantile dependence in the Clayton copula increases the probability of seeing a high maximum underlying, increasing the value of these options at medium and high strikes. For low strike Calls-on-Best, this effect is mitigated by the high lower quantile dependence of the Clayton model, which results in a lower minimum of the distribution of maximum returns[2]. For the Student's t copula, the high levels of upper and lower quantile dependence cause the distribution of the maximum to have lower value, lowering the price of the option. In this case it appears that the empirical data behave most like the Clayton model, having low upper quantile dependence and high lower quantile dependence.

For Call-on-Worst options, the low upper quantile dependence reduces the maximum of the minimum, leading to lower prices at medium and high strikes. Again, this effect is mitigated at low strikes by the higher minimum of the minimum returns. For the Student's t copula, the high levels of upper and lower quantile dependence cause the distribution of the minimum to take higher value, increasing the price of the option. Again, the empirical data behave most like the Clayton model.

Note the close relationship between the prices of Calls-on-Best and Calls-on-Worst, a consequence of the no-arbitrage relationships described above. We can also observe that, modulo Monte Carlo error, each model results in an identical value of implied correlation as a function of K.

[1]The results for these products are a little more difficult to explain than in the previous examples, and the writer apologises to the reader for the linguistic gymnastics that follow.

[2]See what I mean.

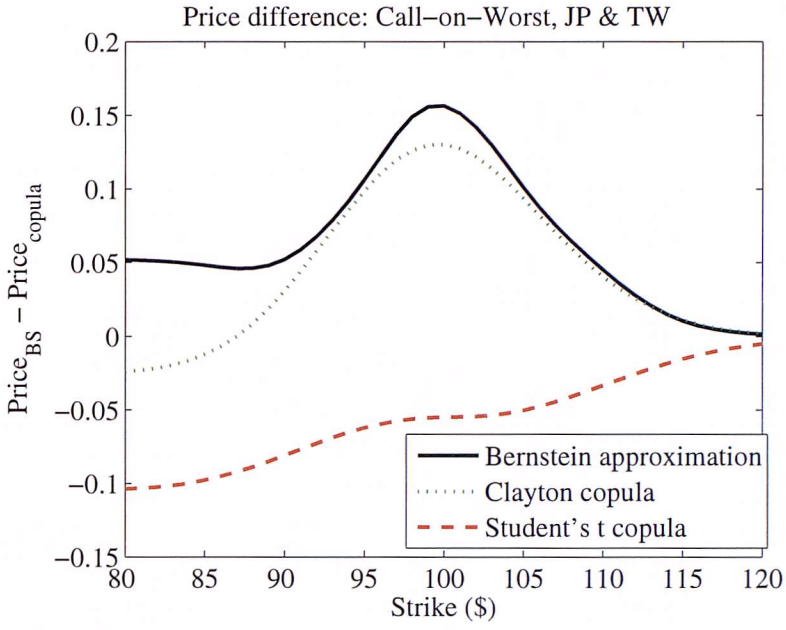


Figure 5.9: Absolute price differences for a Call-on-Worst option

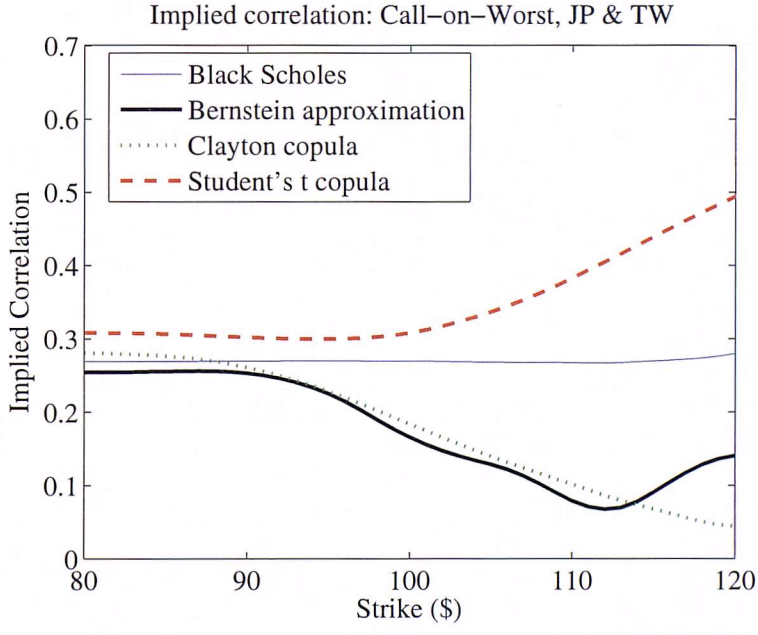


Figure 5.10: Implied correlations for a Call-on-Worst option

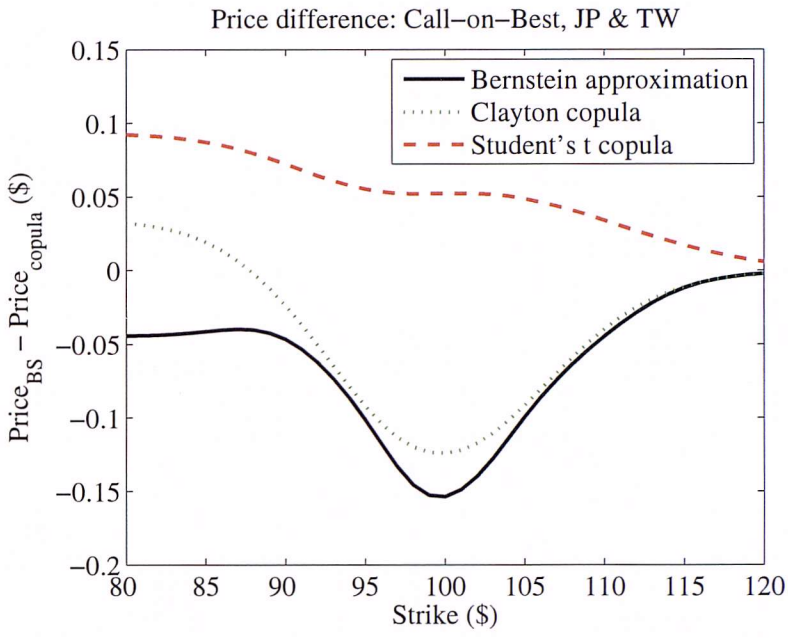


Figure 5.11: Absolute price differences for a Call-on-Best option

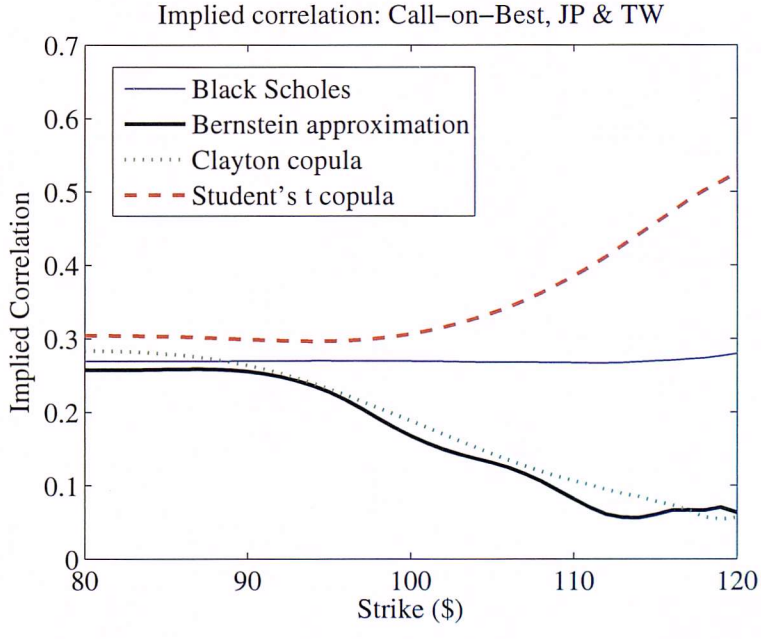


Figure 5.12: Implied correlations for a Call-on-Best option

Chapter 6

Summary

We have reviewed and developed parametric and non-parametric copula techniques, with aim of finding models for the dependence between financial assets that reflect empirical observation and are practically applicable. We began by defining copulas mathematically, and showing some of the key properties of these functions. We described a number of commonly used measures of statistical association, paying particular attention to the concept of 'tail dependence'. It was argued that tail dependence as it is currently defined is of little use to practioners as it is defined asymptotically, and introduced the concept of lower and upper quantile dependence which we suggest is more useful. The practical issues of calibrating a copula model to empirical data and generating random samples from a particular model were also considered in general terms.

Next, we examined several of the parametric copula models frequently cited in the quantitative finance literature. The widely applied Gaussian copula was shown to underestimate the lower quantile dependence of empirical data, while the Student's t copula exhibits greater levels of tail dependence but does not exhibit the asymmetry observed empirically. The Clayton and (rotated) Gumbel copulas do show such asymmetry, but the fact that they have just one parameter makes calibration difficult in multiple dimensions. More sophisticated models like the grouped t and mixture copulas also capture asymmetry effectively, but are again difficult to calibrate easily.

We developed a new copula, that implied by the Gaussian Mixture distribution, which we aimed to show was as tractable as some of the simpler models, while still being able to model empirical data well. We showed that this model does not suffer from the excess of symmetry displayed by the elliptical models, and the number of free parameters are more than sufficient to fit to empirical data. Analytic expressions are derived for τ, ρ_S, λ_L and λ_U for this copula - and by extension for the Gaussian mixture

distribution more generally. We also described one approach to calibrating this model, which is based on pairwise measures of dependence and hence both numerically efficient and robust to poor quality data, and a simple algorithm for generating Monte Carlo samples.

The chapter on non-parametric models began by highlighting some inconsistencies in the existing literature on 'empirical copulas'. We defined the empirical sub-copula, then used Partitions of Unity to extend this to the empirical copula in a rigorous way. To make this possible, the results of [Li 98] and [Kul 99] were generalised to cover the extension of a sub-copula in d dimensions. The Checkerboard and Bernstein Partitions of Unity were examined and compared; by calibrating them to a sample dataset we were able to confirm the result of [San 04] that the Bernstein approximation is much more efficient, and to show that the likelihood cross validation technique is an effective and efficient way of calibrating the Bernstein approximation. We examined the kernel copula proposed by [Fer 03], and saw that it displayed a similar level of accuracy to the Bernstein approximation. By viewing the Gaussian kernel copula as a special case of the GM copula, we were also able to compute τ, ρ_S, λ_L and λ_U for this approach. The first effective simulation algorithms for the Partition of Unity and kernel approximations are developed, and in the former case shown to be as efficient as the best sampling techniques from simplest parametric models.

We concluded by demonstrating how copulas in general and our proposed models in particular can be applied to financial problems. A range of copula models were used in conjunction with the extreme value theory of [Dan 00] to compute the market risk on a portfolio of Asian equity indices. The Gaussian Mixture model performed considerably better in statistical tests than the Gaussian, Student's t or Clayton model. The first systematic investigation of the effect of copulas on the price of basket options showed several interesting results, which were explained with reference to the nature of each copula model. Of particular interest was the effect of pricing with non-parametric model, which was possible for the first time thanks to the simulation algorithms developed in the previous chapter.

There are a number of potential courses for further research on this topic. It would be of interest to compare the prices implied by the various copula models by those quoted in the market for multi-underlying options, but for now the lack of a transparent and liquid market in such products precludes this. The Gaussian Mixture and non-parametric models could also be used to model assets other than equities – indeed we hope that they could be found useful outside of quantitative finance altogether.

Appendix A

Probability Theory

In this appendix we recall some of the basic ideas underlying the study of probability and random variables, and collect some results that will be used in the body of the thesis. For a concise introduction to measure theoretic probability see [Cap 04] or [Gri 01]; a thorough discussion can be found in [Bil 95].

A.1 Elementary Probability Theory

We begin by introducing some of the basic concepts in probability theory which will enable us to state precisely what we mean by a 'random variable'.

Consider some non-empty set Ω, which we will refer to as a *sample space*. $\mathcal{F}$, a collection of subsets of Ω, is called a *σ-field* if it satisfies the following conditions

1. The empty set is a member of $\mathcal{F}$; i.e. $\emptyset \in \mathcal{F}$.
2. $\mathcal{F}$ is closed under countable unions; i.e. if $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.
3. $\mathcal{F}$ is closed under complements; i.e. if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.

A *probability measure* P on $(\Omega, \mathcal{F})$ is a function $P : \mathcal{F} \to [0, 1]$ such that

1. $P(\emptyset) = 0$, $P(\Omega) = 1$.
2. If $A_1, A_2, \ldots$ is a collection of disjoint members of $\mathcal{F}$, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

The triple $(\Omega, \mathcal{F}, P)$ is called a *probability space.* We now have all the tools required to define a random variable, but first let us mention a few results that can be formulated prior to making this definition.

A key concept in multivariate statistics is conditioning: given that some event B has occurred, how does this affect the probability that some other event A occurs? Given some event B with $P(B) > 0$, the *conditional probability* of A given B is

$$P(A|B) = \frac{P(A \cup B)}{P(B)}.$$

From this definition follows an important lemma concerning the partition of a probability space: for any events A and B, with $0 < P(B) < 1$, we have that

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c);$$

and more generally for some partition $B_1, B_2, \ldots, B_n$ of Ω, we have

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i). \tag{A.1}$$

We can combine the above results to get the well known Bayes' theorem: given a partition $A_1, \ldots, A_n$ of Ω, we have that

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}. \tag{A.2}$$

A.2 Random Variables & Distributions

Given a probability space $(\Omega, \mathcal{F}, P)$, a *random variable* X is a function $X : \Omega \to \mathbf{R}$ with the property $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$ for every $x \in \mathbf{R}$. Associated with every random variable is a probability distribution, and the *distribution function* of a random variable X is a function $F : \mathbf{R} \to [0, 1]$ such that

$$F(x) = P\left(\{\omega \in \Omega : X(\omega) \leq x\}\right) \equiv P(X \leq x).$$

We note the following properties of a distribution function which follow from the above definition:

1. $\lim_{x \to -\infty} F(x) = 0, \lim_{x \to \infty} F(x) = 1.$
2. F is non-decreasing, i.e. if $x < y$, then $F(x) \leq F(y)$.
3. F is right continuous, i.e. $F(x + h) \to F(x)$ as $h \searrow 0$.

We note that the above properties are in fact sufficient to define F as a distribution function. Two random variables X_1 and X_2 are said to be equal in distribution if and only if $P(X_1 \leq x) = P(X_2 \leq x)$ for all x, and we write $X_1 \doteq X_2$

The *quasi-inverse* of a distribution function F is a function $F^{(-1)} : [0,1] \rightarrow \mathbf{R}$ such that

$$F^{(-1)}(u) = \inf\{x : F(x) \geq u\}. \tag{A.3}$$

For F strictly increasing the quasi-inverse is equivalent to the usual inverse, also known as a *quantile function.* The *survival function* of random variable with distribution function F is the function $\overline{F} : \mathbf{R} \rightarrow [0,1]$ such that

$$\bar{F}(x) = P(X > x) = 1 - F(x).$$

A random variable is called *continuous* if its distribution function can be written

$$F(x) = \int_{-\infty}^{x} f(u)du$$

for some integrable function $f : \mathbf{R} \rightarrow [0, \infty)$, where f is called a *density function.* If F if is differentiable at x then we have that

$$f(x) = \frac{d}{dx}F(x).$$

The distribution function of a continuous random variable is absolutely continuous. The *expectation* of a continuous random variable can be defined via its density function as

$$E[X] = \int xf(x)dx,$$

and more generally we can define the i-th moment of X to be

$$E[X^i] = \int x^i f(x)dx.$$

A collection of random variables $\mathbf{X} = (X_1, \ldots, X_d)$ defined with respect to the same probability space $(\Omega, \mathcal{F}, P)$ is called a *random vector.* A random vector has associated with it a *joint distribution function* $F_{\mathbf{X}} : \mathbf{R}^d \rightarrow [0,1]$ given by

$$F_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{x} \leq \mathbf{X})$$

for $\mathbf{x} \in \mathbf{R}^d$. Note that we write $\mathbf{x} < \mathbf{y}$ if and only if $x_i \leq y_i \, \forall \, i = 1, \ldots, d$. We can write down the defining properties of a joint distribution function in a similar fashion to those of an ordinary distribution function:

1. $\lim_{x_i \rightarrow -\infty} F_{\mathbf{X}}(\mathbf{x}) = 0$, $\lim_{x_i \rightarrow \infty} F_{\mathbf{X}}(\mathbf{x}) = 1$ for all x_i.

2. F is d-increasing, i.e. if $\mathbf{x} < \mathbf{y}$, then $F_{\mathbf{X}}(\mathbf{x}) < F_{\mathbf{X}}(\mathbf{y})$.

The i-th*marginal distribution function* of $F_{\mathbf{X}}(\mathbf{x})$ is defined as

$$F_i(x_i) \equiv P[X_i \leq x_i] = F_{\mathbf{X}}(\infty, \ldots, x_i, \ldots, \infty)$$

The *multivariate survival function* of a random vector is the function

$$\bar{F}(\mathbf{x}) = P[\mathbf{X} > \mathbf{x}].$$

For the bivariate case we have that

$$\bar{F}(x_1, x_2) = 1 - F_1(x_1) - F_2(x_2) + F(x_1, x_2)$$

A random vector is continuous if we can write

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(u_1, \ldots, u_d) du_1 \cdots du_d$$

for some integrable function $f_{\mathbf{X}} : \mathbf{R}^d \to [0, \infty)$, where f is called a *joint density function.* If $F_{\mathbf{X}}$ is sufficiently differentiable at $\mathbf{x}$ then we can write

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^d}{\partial x_1 \cdots \partial x_d} F_{\mathbf{X}}(\mathbf{x}).$$

For a continuous bivariate distribution function $F_{\mathbf{X}}$ the *conditional distribution function* of X_2 given that $X_1 = x_1$ is written $F_{X_2|X_1}(x_2|x_1)$ and is given by

$$F_{X_2|X_1}(x_2|x_1) = \int_{-\infty}^{x_2} \frac{f_{\mathbf{X}}(x_1, t)}{f_{X_1}(x_1)} dt,$$

where $f_{X_1}(x_1)$ denotes the marginal density function of X_1

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2$$

The density function of $F_{X_2|X_1}$ is called the *conditional density function* and is given by

$$f_{X_2|X_1}(x_2|x_1) = \frac{f(x_1, x_2)}{f_{X_1}(x_1)}.$$

The definitions of conditional distribution function, conditional density function and marginal density function extend to the multivariate case in the obvious fashion, for conciseness of exposition we omit them here.

A.3 Elliptical Distributions

We now give some examples of elliptical distributions, a particular class of multivariate distribution which is be referred to extensively in the thesis. For a thorough discussion of elliptical distributions, see [Fan 87]. A random vector $\mathbf{x}$ has a *spherical distribution* if and only if, for any orthogonal $\Gamma \in R^{d \times d}$,

$$\Gamma \mathbf{x} \doteq \mathbf{x}.$$

A random vector $\mathbf{x}$ is said to have an *elliptical distribution* with parameters $\mu \in R^d$ and $\Sigma \in R^{d \times d}$ if and only if

$$\mathbf{x} \doteq \mu + A'\mathbf{y}$$

for some spherically distributed $\mathbf{y}$ and $AA' = \Sigma$.

Let us now focus on two examples of elliptical distributions. A random vector $\mathbf{X}$ has a *multivariate normal* or *multivariate Gaussian* distribution if and only if it has density function

$$\phi_d(x_1, \ldots, x_d) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right),$$

and distribution function

$$\Phi_d(x_1, \ldots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{u}-\mu)'\Sigma^{-1}(\mathbf{u}-\mu)\right) d\mathbf{u}.$$

The parameters μ and Σ are referred to as the mean vector and covariance matrix respectively. We can decompose the covariance matrix into a dispersion matrix R and variance vector σ as

$$\Sigma_{ij} = \sigma_i R_{ij} \sigma_j \quad \text{for all } i, j = 1, \ldots, d.$$

A random vector $\mathbf{X}$ has a *multivariate Student's t* distribution if and only if it has density function

$$t_d(x_1, \ldots, x_d) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)|\Sigma|^{-1/2}}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{d/2}} \left(1 + \frac{1}{\nu}\mathbf{x}'\Sigma^{-1}\mathbf{x}\right)^{-(\nu+d)/2},$$

and distribution function

$$T_d(x_1, \ldots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} \frac{\Gamma\left(\frac{\nu+d}{2}\right)|\Sigma|^{-1/2}}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{d/2}} \left(1 + \frac{1}{\nu}\mathbf{u}'\Sigma^{-1}\mathbf{u}\right)^{-(\nu+d)/2} d\mathbf{u}.$$

The covariance matrix Σ can again be decomposed into a variance vector and dispersion matrix, while the parameter $\nu \in [0, \infty)$ is referred to as the number of degrees of freedom.

A.4 Random Processes

Geometric Brownian Motion, the model for the evolution of asset prices used by [Bla 73], is based on the *Standard Wiener Process* $W = \{W(t) : t \geq 0\}$, a real-valued process such that

- $W(0) = 0$
- W has independent increments
- The increments of W are are distributed as $W(s+t) - W(t) \sim \Phi(0, t^2)$ for all $s, t \geq 0$,
- The sample paths of W are continuous

This object displays a number of interesting properties, see for example [Gri 01] §13. We call the process

$$dX(t) = a(t, X(t)) + b(t, X(t))dW(t)$$

Brownian Motion, and the process

$$dX(t) = a(t, X(t))X(t) + b(t, X(t))X(t)dW(t)$$

Geometric Brownian Motion. Setting $a(t, X(t) = \mu$ and $b(t, X(t)) = \sigma$ gives the Black-Scholes model for asset prices.

Appendix B

Summary Statistics for Sample Data

	HK	JP	KO	SG	TH	TW
Start Date	01/03/82	05/02/82	04/10/84	08/01/85	06/01/82	03/09/81
End Date	01/06/06	01/06/06	01/06/06	01/06/06	01/06/06	01/06/06
Sample Length	5999	5996	5300	5356	5998	5997
Mean	0.57×10^{-3}	0.20×10^{-3}	0.59×10^{-3}	0.35×10^{-3}	0.44×10^{-3}	0.62×10^{-3}
Variance	0.29×10^{-3}	0.18×10^{-3}	0.32×10^{-3}	0.20×10^{-3}	0.25×10^{-3}	0.38×10^{-3}
Skewness	-1.59	0.04	0.10	-0.97	0.25	0.34
Kurtosis	38.2	10.3	7.0	45.8	9.6	11.8
Minimum	-0.33	-0.15	-0.12	-0.25	-0.10	-0.18
1st percentile	-0.046	-0.035	-0.049	-0.034	-0.043	-0.057
5th percentile	-0.024	-0.021	-0.027	-0.018	-0.023	-0.029
Median	0.68×10^{-3}	0.37×10^{-3}	0.44×10^{-3}	0.26×10^{-3}	0.32×10^{-3}	0.39×10^{-3}
95tht percentile	0.025	0.020	0.028	0.019	0.024	0.030
99th percentile	0.044	0.036	0.054	0.036	0.048	0.055
Maximum	0.19	0.13	0.11	0.18	0.12	0.22
Lower tail index	3.18	2.70	2.31	2.94	2.17	2.52
Upper tail index	2.80	3.52	3.20	3.48	2.84	2.92

Table B.1: Univariate summary statistics for daily returns of Asian equity indexes

	HK	JP	KO	SG	TH	TW
HK	1	0.30	0.24	0.51	0.28	0.16
JP	0.30	1	0.22	0.28	0.18	0.18
KO	0.24	0.22	1	0.22	0.22	0.17
SG	0.51	0.28	0.22	1	0.37	0.19
TH	0.28	0.18	0.22	0.37	1	0.16
TW	0.16	0.18	0.17	0.19	0.16	1

Table B.2: Matrix of Linear Correlations

	HK	JP	KO	SG	TH	TW
HK	1	0.19	0.14	0.30	0.15	0.11
JP	0.19	1	0.13	0.19	0.09	0.12
KO	0.14	0.13	1	0.14	0.12	0.12
SG	0.30	0.19	0.14	1	0.21	0.14
TH	0.15	0.09	0.12	0.21	1	0.10
TW	0.11	0.12	0.12	0.14	0.10	1

Table B.3: Matrix of Kendall's Tau

	HK	JP	KO	SG	TH	TW
HK	1	0.28	0.21	0.42	0.22	0.16
JP	0.28	1	0.19	0.28	0.13	0.17
KO	0.21	0.19	1	0.21	0.18	0.18
SG	0.42	0.28	0.21	1	0.30	0.20
TH	0.22	0.13	0.18	0.30	1	0.14
TW	0.16	0.17	0.18	0.20	0.14	1

Table B.4: Matrix of Spearman's Rho

	HK	JP	KO	SG	TH	TW
HK	1	0.21	0.22	0.41	0.25	0.11
JP	0.21	1	0.19	0.24	0.19	0.16
KO	0.22	0.19	1	0.24	0.19	0.15
SG	0.41	0.24	0.24	1	0.36	0.14
TH	0.25	0.19	0.19	0.36	1	0.17
TW	0.11	0.16	0.15	0.14	0.17	1

Table B.5: Matrix of Lower Quantile Dependence at $u = 0.05$

	HK	JP	KO	SG	TH	TW
HK	1	0.19	0.19	0.35	0.23	0.12
JP	0.19	1	0.20	0.20	0.14	0.16
KO	0.19	0.20	1	0.20	0.20	0.15
SG	0.35	0.20	0.20	1	0.26	0.15
TH	0.23	0.14	0.20	0.26	1	0.12
TW	0.12	0.16	0.15	0.15	0.12	1

Table B.6: Matrix of Upper Quantile Dependence at $u = 0.95$

Bibliography

[Aït 98] Aït-Sahalia, Y. & Lo, A., 'Nonparametric Estimation of State-Price Densities Implict in Finacial Asset Prices', *Journal of Finance*, 53:499-547.

[Ang 02] Ang, A. & Chen, J., 2002, 'Asymmetric Correlations of Equity Portfolios', *Journal of Financial Economics*, 63:443-494.

[Art 99] Artzner, P., Delbaen, F., Eber, J.-M. & Heath, D., 1999, 'Coherent Measures of Risk', *Mathematical Finance*, 9:203-228.

[Bab 02] Babu, G., Canty, A. & Chaubey, Y., 2002, 'Application of Bernstein Polynomials for Smooth Estimation of a Distribution and Density Function', *Journal of Statistical Planning and Inference*, 105:377-392.

[Bas 05] Basel Committee on Banking Supervision, 2005, *International Convergence of Capital Measurement and Capital Standards: A Revised Framework*, Bank for International Settlements.

[Ber 12] Bernstein, S., 1912, 'Démonstration du théorème de Weierstrass fondée sur le calcul des probabilities', *Communications of the Kharkov Mathematical Society*, 13:1-2.

[Bil 95] Billingsley, P., 1995, *Probablility and Measure*, Wiley.

[Bla 73] Black, F. & Scholes, M., 1973, 'The Pricing of Options and Corporate Liabilities', *Journal of Political Economy*, 81:637-659.

[Bla 88] Black, F., 1988, 'The Holes in Black-Scholes', *Risk*, 1/4:30-32.

[But 53] Butzer, P., 1953, 'On Two Dimensional Bernstein Polynomials', *Canadian Journal of Mathematics*, 5:107-113.

[Cap 04] Capinski, M. & Kopp, P., 2004, *Measure, Integral and Probablility*, Springer.

[Che 02] Cherubini, U. & Luciano, E., 2002, 'Bivariate Option pricing with Copulas', *Applied Mathematical Finance*, 9:69-86.

[Che 04] Cherubini, U., Luciano, E. & Vecchiato, W., 2004, *Copula Methods in Finance*, Wiley Finance.

[Col 99] Coles, S., Heffernan J. & Twan, J., 1999, 'Dependence Measures for Extreme Value Analyses', *Extremes*, 2:339-365.

[Dan 00] Danielsson, J. & de Vries, C., 2000, 'Value-at-Risk and Extreme Returns', *Annals d'Economie et de Statistique*, 60:239-269.

[Dar 92] Darsow, W., Hguyen, B. & Olsen, E., 1992, 'Copulas and Markov Processes', *Illinois Journal of Mathematics*, 36:600-642.

[Dat 02] Datta, G., 2002, 'Symmetry of Finite Mixtures of Normal Distributions', working paper, University of Georgia.

[Deh 80] Deheuvels, P., 1980, 'Non Parametric Tests of Independence', *Springer Lecture Notes in Statistics*, 820:95-107.

[Dem 05] Demarta, S. & McNeil, A., 2005, 'The t Copula and Related Copulas', *International Statistical Review* 73, 111-129.

[Dev 85] Devroye, L. & Györfi, L., 1985, *Nonparametric Density Estimation: The L_1 view*, John Wiley & Sons.

[Dev 86] Devroye, L., 1986, *Non-Uniform Random Variate Generation*, Springer-Verlag.

[Dow 02] Dowd, K., 2002, *An Introduction to Market Risk Measurement*, John Wiley & Sons.

[Dur 00a] Durrleman, V., Nikeghbali, A. & Roncalli, T., 2000, 'Copulas Approximation and New Families', working paper, Crédit Lyonnais.

[Dur 00b] Durrleman, V., Nikeghbali, A. & Roncalli, T., 2000, 'Which Copula is the Right One?', working paper, Crédit Lyonnais.

[Emb 01] Embrechts, P., Lindskog, F. & McNeil, A., 2001, 'Modelling Dependence with Copulas and Applications to Risk Management', working paper, ETH Zurich.

[Efr 79] Efron, B., 1979, 'Bootstrap Methods: Another Look at the Jackknife', *The Annals of Statistics*, 7:1-26.

[Fam 65] Fama, E., 1965, 'The Behavior of Stock Market Prices', *Journal of Business*, 38:34-105.

[Fan 87] Fang, K., Kotz, S. & Ng, K., 1987, *Symmetric Multivariate and Related Distribution*, Chapman & Hall.

[Fen 03] Fengler, M. & Schwender, P., 2003, 'Correlation Risk Premia for Multi-Asset Equity Options', working paper, Humboldt University of Berlin.

[Fer 03] Fermanian J.-D. & Scaillet, O., 2003, 'Nonparametric Estimation of Copulas for Time Series', *Journal of Risk*, 5:25-54.

[Fer 04] Fermanian, J.-D., Radulović, D. & Wegkamp, M., 2004, 'Weak Convergence of Empirical Copula Processes', *Bernoulli*, 10:847-860.

[Fra 05] Frahm, G., Junker, M. & Schmidt, R., 2005, 'Estimating the tail-dependence coefficient', *Insurance: Mathematics and Economics*, 37:80-100.

[Gen 93] Genest, C. & Rivest, L., 1993, 'Statistical Inference Procedures for Bivariate Archimedean Copulas', *Journal of the American Statistical Association*, 88:1034-1043.

[Geo 01] Georges, P., Lamy, A., Nicolas, E., Quibel, G. & Roncalli, T., 2001, 'Multivariate Survival Modelling: A Unified Approach with Copulas', working paper, Crédit Lyonnais.

[Goo 05] van den Goorbergh, R., Genest, C. & Werker, B., 2005, 'Bivariate Option Pricing using Dynamic Copula Models', *Insurance: Mathematics and Economics*, 37:101-114.

[Gri 01] Grimmett, G. & Stirzaker, D., 2001, *Probablility and Random Processes*, Oxford University Press.

[Hil 75] Hill, B., 1975, 'A Simple General Approach to Inference about the Tails of a Distribution', *Annals of Statistics*, 25:1163-1173.

[Hör 00] Hörmann, W. & Leydold, J., 2000, 'Automatic Random Variate Generation for Simulation Input', In: Joines, J., Barton, R., Kang, K. & Fishwick, P. (Eds.), *Proceedings of the 2000 Winter Simulation Conference*, pp. 675-682.

[Hu 03] Hu, L., 2003, 'Dependence Patterns across Financial Markets: a Mixed Copula Approach', working paper, Yale University.

[Hul 03] Hull, J., 2003 , *Options, Futures and Other Derivatives, 5th Edition*, Prentice Hall.

[Joe 97] Joe, H., 1997, *Multivariate Models and Dependence Concepts*, Chapman & Hall.

[Joh 87] Johnson, H., 1987, 'Options on the Maximum and Minimum of Several Assets' *Journal of Financial and Quantitative Analysis*, 22:277-283.

[Kir 95] Kirk, E. & Aron, J., 1995, *Managing Energy Price Risk*, Risk Books.

[Kul 99] Kulpa, T., 1999, 'On Approximation of Copulas', *International Journal of Mathematics and Mathematical Sciences*, 22:259-269.

[Li 97] Li, X., Mikusiński, P., Sherwood, H. & Taylor, M., 1997, 'On Approximation of Copulas', In: Benes, V. & Stepán, J. (Eds.), *Distributions with Given Marginals and Moment Problems*, pp 107-116, Kluwer.

[Li 98] Li, X., Mikusiński, P. & Taylor, M., 1998, 'Strong Approximation of Copulas', *Journal of Mathematical Analysis and Applications*, 225:608-623.

[Li 00] Li, D., 2000, 'On Default Correlation: A Copula Function Approach'. *Journal of Fixed Income*, 9:43-54.

[Lin 03] Lindskog, F., McNeil, A. & Schmock, U., 2001, 'Kendall's Tau for Elliptical Distributions', In: Bol, Nakhaeizadeh, Rachev, Ridder & Vollmer, (Eds.), *Distributions with Given Marginals and Moment Problems*, Physica-Verlag .

[Lon 01] Longin, F. & Solnik, B., 2001, 'Extreme Correlation of International Equity Markets', *Journal of Finance*, 56:649-679.

[Luc 05] Luciano, E. & Schoutens, W., 2005 'A Multivariate Jump-Driven Financial Asset Model', working paper, Université Catholique de Louvain.

[Mar 78] Margrabe, W., 1978, 'The Value of an Option to Exchange One Asset for Another', *Journal of Finance*, 33:177-186.

[Man 63] Mandelbrot, B., 1963, 'The Variation of Certain Speculative Prices', *Journal of Business*, 36:143-160.

[Mar 52] Markowitz, H., 1952, 'Portfolio Selection', *Journal of Finance*, 7:77-91.

[Mar 88] Marshall, A. & Olkin, I., 1988, 'Families of Multivariate Distribution', *Journal of the American Statistical Associtaion*, 83:834-841.

[Mas 02] Mashal, R. & Zeevi, A., 2002, 'Beyond Correlation: Extreme Co-movements between Financial Assets', working paper, Columbia University.

[McL 00] McLachlan, G. & Peel, D., 2000, *Finite Mixture Models*, Wiley.

[Mel 06] Melchiori, M., 2006, 'Tools for Sampling Multivariate Archimdean Copulas', working paper, Universidad Nacional del Litoral.

[Mus 05] Musiela, M. & Rutkowski, M., 2005, *Martingale Methods in Finacial Modelling*, Springer.

[Nel 99] Nelsen, R., 1999, *An Introduction to Copulas*, Springer-Verlag.

[Pre 92] Press, W., Teukolsky, S., Vetterling, W. & Flannery, B., 1992, *Numerical Recipes in C: The Art of Scientific Computing, 2nd Ed.*, Cambridge University Press.

[Qu 06] Qu, D., 2006, 'Pricing Basket Options with Skew', *Wilmott magazine*, ?? 2006:58-64.

[Rac 03] Rachev, S., 2003, *Handbook of Heavy Tailed Distributions in Finance*, North-Holland.

[Rap 01] Rapuch, G. & Roncalli, T., 2001, 'Some Remarks on Two-Asset Option Pricing and Stochastic Dependence of Asset Prices', working paper, Crédit Lyonnais.

[Reb 99] Rebonato, R. & Jäckel, P., 'The Most General Methodology for Creating a Valid Corrleation Matrix for Risk Management and Option Pricing Purposes', *Journal of Risk*, 2:.

[Rit 90] Ritchey, R., 'Call Option Valuation for Discrete Normal Mixtures', *Journal of Financial Research*, 13:285-295.

[Ros 03] Rosenberg, J., 2003, 'Nonparametric Pricing of Multivariate Contingent Claims', *Journal of Derivatives*, 10:9-26.

[San 04] Sancetta, A. & Satchell, S., 2004, 'The Bernstein Copula and its Applications to Modeling and Approximations of Multivariate Distributions', *Econometric Theory*, 20:535-562.

[San 05] Sancetta, A., 2005, 'Copula Based Monte Carlo Integration in Financial Problems', working paper, Cambridge University.

[Sca 05] Scaillet, O., 2005, 'A Kolmogorov-Smirnov type test for Positive Quadrant Dependence', *The Canadian Journal of Statistics*, 33: in press.

[Sch 03a] Schmidt, R., 2003, 'Tail Dependence', In: Härdle, W., Cizek, P. & Weron, R. (Eds.), *Statistical Tools in Finance and Insurance*, Springer-Verlag.

[Sch 03b] Schönbucher, P., 2003, *Credit Derivatives Pricing Models*, John Wiley & Sons.

[Sil 86] Silverman, B., 1986, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall.

[Skl 59] Sklar, A., 1959, 'Fonctions de répartition à n dimensions et leurs marges', *Publications de l'Institut de Statistique de l'Unviversité de Paris*, 8:229-231.

[Skl 96] Sklar, A., 1996, 'Random Variables, Distribution Functions and Copulas – a personal look backward and forward', In: Rüschendorff, L., Schweizer, B. & Taylor, M., *Distributions with Fixed Marginals and Related Topics*, pp 1-14, Institute of Mathematical Statistics.

[Stu 82] Stulz, R., 1982, 'Options on the Minimum or the Maximum of Two Risky Assets: Analysis and Applications', *Journal of Financial Economics*, 10:161-185.

[Ten 94] Tenbusch, A., 1994, 'Two-Dimensional Bernstein Polynomial Density Estimators, *Metrika*, 41:233-253.

[Vit 75] Vitale, R., 1975, 'A Bernstein Polynomial Approach to Density Function Estimation', In: Puri, M. (Ed.), *Statistical Inference and Related Topics*, Vol. 2, pp 87-89, Academic Press.

[Wan 95] Wand, M. & Jones, M., 1995, *Kernel Smoothing*, Chapman & Hall.

[Wes 05] West, G., 2005, 'Better Approximations to Cumulative Normal Functions', *Wilmott Magazine*, May 2005, pp 70-76.