Pricing multi-asset financial products with tail dependence using copulas

Master's Thesis

J.P. de Kort

Delft University of Technology
Delft Institute for Applied Mathematics

and

ABN AMRO Bank N.V.
Product Analysis

Thesis Committee
Prof. dr. ir. C. Vuik
Dr. ir. C.W. Oosterlee
Dr. R.J. Fokkink
Dr. A.C.P. Gee
Dr. J. Smith

October 1, 2007

Contents

1 Introduction

Suppose we want to price a 'best-of' option on corn and wheat, that is, a contract that allows us to buy a certain amount of wheat *or* corn for a predetermined price ('strike') on a predetermined date ('maturity'). The value of such a contract will only decrease if both the price of corn and the price of wheat move down. In pricing a best-of option, the probability of simultaneous extreme movements — usually called 'tail dependence' — thus is of great importance. From a Gaussian perspective, tail dependence is observed as 'increasing correlation' as the underlying quantities simultaneously move towards extremes. This thesis is concerned with the question how tail dependence can be incorporated in pricing models and how it affects prices and hedging of financial contracts.

In option pricing it is common to model the dependence structure between assets using a Gaussian copula. Copulas are a way of isolating dependence between random variables (such as asset prices) from their marginal distributions. In section 5.2.1 it will be shown that the Gaussian copula does not have tail dependence. This may cast some doubt on the appropriateness of this model for underlyings that have a high probability of joint extreme price movements.

Malevergne and Sornette [1] show that a Gaussian copula may indeed not always be a feasible choice. They succeed in rejecting the hypothesis of the dependence between a number of metals traded on the London Metal Exchange being described by a Gaussian copula.

For the univariate case similar problems have been dealt with earlier: the classic Black-Scholes model assumed a normal distribution for daily increments of the underlyings underestimating the probability of extreme (univariate) price changes. This is usually solved by using a parametrization of equivalent normal volatilities, i.e. the volatilities that lead to the correct market prices when used in the Black-Scholes model instead of one constant number. Due to the typical shape of such parametrizations the problem of underestimation of univariate tails is usually referred to as 'volatility smile'. Tail dependence similarly leads to a 'correlation skew' in the implied correlation surface.

In our adjusted model, we want to be able to account for both tail dependence and — to be consistent with the univariate case — volatility smile. Copulas are a convenient tool in modelling dependence since the marginal distributions of the underlyings can be specified independently. However, we will have to look at copulas other than the Gaussian.

The first part of this thesis gives an overview of the theory of copulas and dependence. Sections 2 and 3 explain what copulas are and how they relate to multivariate distribution functions. In Section 4 it is described what kind of dependence is captured by copulas. This, among other things, includes measures of concordance like Kendall's tau and Spearman's rho. Next, Section 5 summarizes the properties of a number of well-known parametric families of copulas.

The second part of the thesis describes how copulas can be used in pricing multi-asset financial derivatives. A pricing model is outlined in Section 6. This model relies on Monte Carlo methods to calculate option prices. For this we need to generate samples from Archimedean copulas. This is the topic of Section 7.

Before copulas can be used in a pricing model, they have to be calibrated. This is discussed in Section 8. Results can be found in Section 9 followed by a conclusion in Section 10.

Appendix A provides a brief introduction to derivatives pricing. Appendix B describes some alternative models for univariate asset price distributions.

1.1 Evidence of tail dependence

How can we recognise the presence of tail dependence in pairs of financial assets? Informally speaking, tail dependence expresses the probability of a random variable taking extreme values conditional on another random variable taking extremes — a formal definition can be found in Section 4.4. For two random variables X, Y with respective distribution functions F, G the definition reads

$$\text{Lower tail dependence coefficient} = \lim_{u \downarrow 0} \frac{\mathbb{P}[F(X) < u,\, G(Y) < u]}{\mathbb{P}[G(Y) < u]}, \tag{1.1}$$

$$\text{Upper tail dependence coefficient} = \lim_{u \uparrow 1} \frac{\mathbb{P}[F(X) > u, G(Y) > u]}{\mathbb{P}[G(Y) > u]}. \tag{1.2}$$

Given a set of historical observations from (X, Y) consisting of the pairs (x_i, y_i), $1 \leq i \leq n$, for fixed u the probabilities in (1.1) can be approximated by their empirical counterparts:

$$\mathbb{P}[F(Y) < u, G(Y) < u] = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(F^{\text{emp}}(x_i) < u,\, G^{\text{emp}}(y_i) < u),$$

$$\mathbb{P}[G(Y) < u] = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(G^{\text{emp}}(y_i) < u),$$

where

$$F^{\text{emp}}(u) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(x_i < u), \qquad G^{\text{emp}}(u) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(y_i < u).$$

A similar approach may be used to approximate (1.2).

Figures 1.1—1.6 show estimates of lower and upper tail dependence in historical asset returns. The solid line represents the empirical approximation to the conditional probabilities (1.1) and (1.2) for fixed u. The tail dependence coefficients are defined as the limit of these probabilities as $u \to 0$ (lower) or $u \to 1$ (upper). If the solid line tends to zero in this limit (Figure 1.4) this means absence of tail dependence. If on the other hand the line does not tend to zero it may indicate the presence of tail dependence (Figures 1.1, 1.2, 1.3, 1.5, 1.6). The dashed lines represent conditional probabilities (1.1) and (1.2) for different copulas calibrated to the data sets. The Clayton copula is seen to have lower tail dependence (limit tends to 0.6 in Figure 1.6) and the Gumbel copula exhibits upper tail dependence (limit tends to 0.1 in Figure 1.2).

A more extensive overview is presented in Table 1.1. It shows that tail dependence is much more profound in daily returns than in price levels.

1.2 Scope of the project

In this thesis we will look at specific contracts over specific underlyings. The choice of underlyings is inspired by our observations in Section 1.1:

Table 1.1: *Empirical evidence for tail dependence in pairs of financial assets (++ = clear empirical evidence of tail dependence, + = possibly tail dependent, − = unclear, −− = no tail dependence).*

	Correlation	Tail dependence Lower	Upper
Daily returns			
Hangseng, SP500	0.3	−	+
Nikkei, SP500	0.3	−−	++
Corn, wheat	0.6	++	−
Nikkei, Hangseng	0.19	−−	+
Oil, gas	−0.04	−−	−−
USD 5Y swap, SP500 futures	0.14	+	++
Copper, nickel	0.48	−−	−−
Gold, copper	0.13	++	−−
Nickel, gold	0.08	−−	+
Price levels			
Corn, wheat	0.87	−−	−
Hangseng, SP500	0.94	−−	−
Nikkei, SP500	0.34	−	−−
USD 5Y swap, SP500 futures	0.54	−−	−
Nikkei, Hangseng	0.41	−−	−

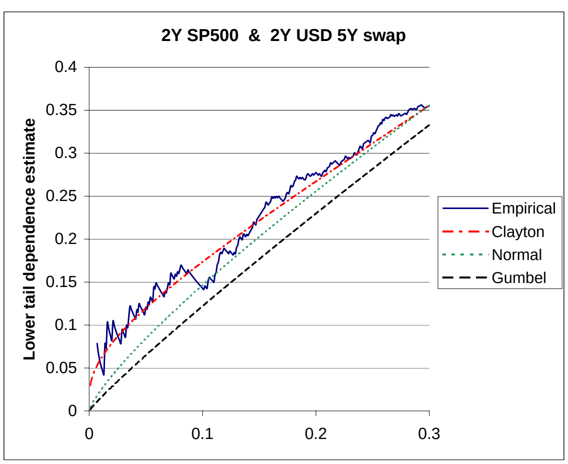


Figure 1.1: *Estimated lower tail dependence coefficient for 2-year SP500 and USD swap futures.*

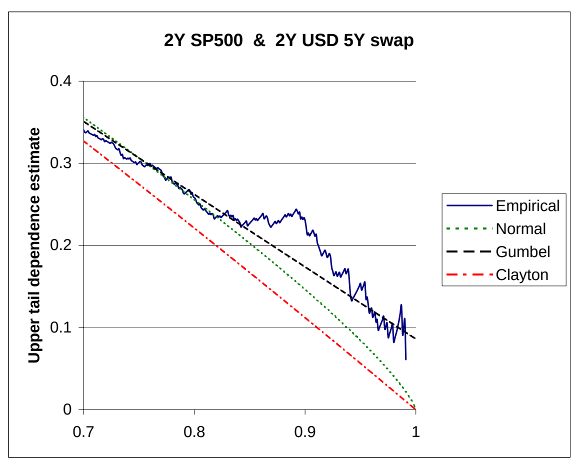
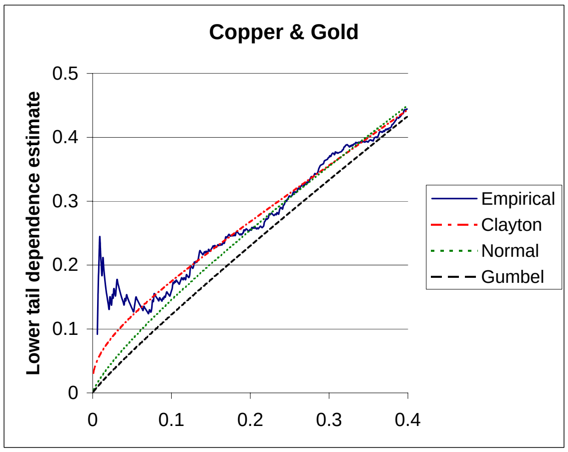


Figure 1.2: *Estimated upper tail dependence coefficient for 2-year SP500 and USD swap futures.*



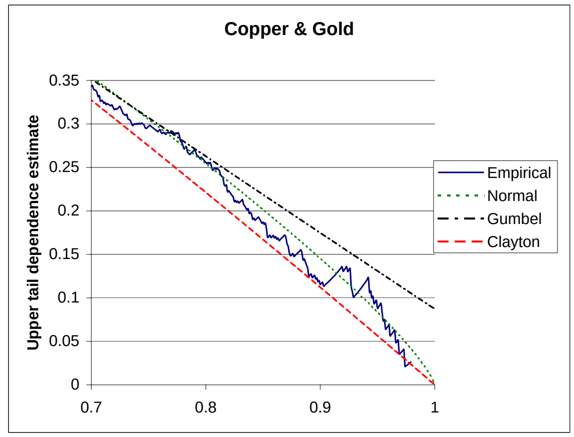
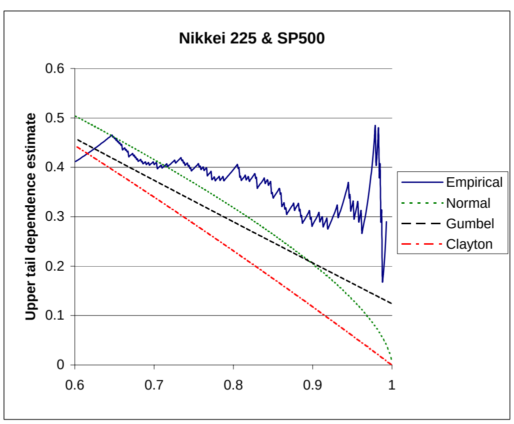


Figure 1.3: *Estimated lower tail dependence coefficient for copper and gold.*

Figure 1.4: *Estimated lower tail dependence coefficient for copper and gold.*



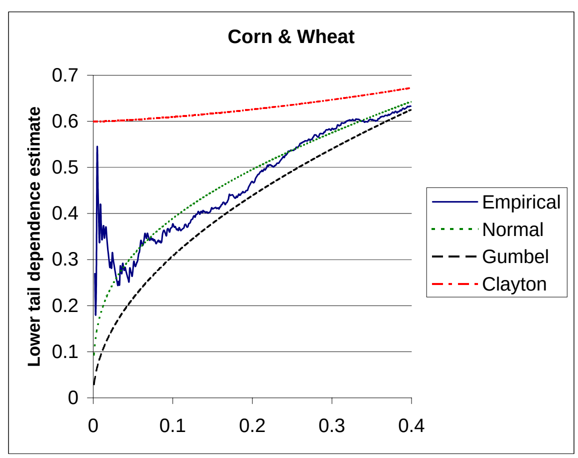


Figure 1.5: *Estimated upper tail dependence coefficient for Nikkei 225 and SP500*

Figure 1.6: *Estimated lower tail dependence coefficient for 1-year corn and wheat futures*

- Nikkei 225 and SP500, August 2001 until August 2007,
- Gold and copper, March 2000 until March 2007,
- Corn and wheat 1-year futures, March 2000 until June 2006, traded on CBOT,
- 2-year futures on SP500 and USD 5-year swaps, March 2000 until December 2006 .

We will focus on European bivariate payoff structures with a single maturity, that is, contracts whose payoff depends on two simultaneous observations (one from each underlying) and the payment is made without delay. The choice of contracts is based on concerns in industry about possible sensitivity to tail dependence:

$$\textbf{Best-of returns} = \max\left(0, \max\left(\frac{S_1(T)}{S_1(0)}, \frac{S_2(T)}{S_2(0)}\right) - 1\right), \tag{1.3}$$

$$\textbf{Worst-of returns} = \max\left(0, \min\left(\frac{S_1(T)}{S_1(0)}, \frac{S_2(T)}{S_2(0)}\right) - 1\right), \tag{1.4}$$

$$\textbf{Spread on returns} = \max\left(0, \frac{S_1(T)}{S_1(0)} - \frac{S_2(T)}{S_2(0)}\right), \tag{1.5}$$

$$\textbf{At-the-money spread} = \max\Big(0, S_1(T) - S_2(T) - S_1(0) + S_2(0)\Big). \tag{1.6}$$

Note that the above payoff structures will be most susceptible to tail dependence if they act on returns or if they are at-the-money, i.e. the strike is chosen such that the value of the option initially is zero. The reason is that prices of different underlyings can be an order of magnitude apart, so that the minimum or maximum of the levels $S_1(T), S_2(T)$ will always stem from the same underlying.

2 Bivariate copulas

This section introduces copulas and describes how they relate to multivariate distributions (Sklar's theorem, Section 2.1). Section 2.2 discusses maximal and minimal bounds for copula functions. In the bivariate case these bounds turn out to be copulas themselves (the Fréchet-Hoeffding copulas). It is further explained what it means for a multivariate distribution if its copula is maximal or minimal. Finally, in Section 2.3, survival copulas will be defined which are essentially a mirrored version of the original copula. In particular, they have the useful property that upper and lower tail properties are interchanged.

The extended real line $\mathcal{R} \cup \{-\infty, +\infty\}$ is denoted by $\overline{\mathcal{R}}$.

Definition 2.1 *Let $S_1, S_2 \subset \overline{\mathcal{R}}$ be nonempty sets and let H be a $S_1 \times S_2 \to \mathcal{R}$ function. The **H-volume** of $B = [x_1, x_2] \times [y_1, y_2]$ where $x_1 \leq x_2$ and $y_1 \leq y_2$ is defined to be*

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1).$$

*H is **2-increasing** if $V_H(B) \geq 0$ for all $B \subset S_1 \times S_2$.*

Definition 2.2 *Suppose $b_1 = \max S_1$ and $b_2 = \max S_2$ exist. Then the **margins** F and G of H are given by*

$$F : S_1 \to \mathcal{R}, \quad F(x) = H(x, b_2),$$
$$G : S_2 \to \mathcal{R}, \quad G(y) = H(b_1, y).$$

Note that b_1 and b_2 can possibly be $+\infty$.

Definition 2.3 *Suppose also $a_1 = \min S_1$ and $a_2 = \min S_2$ exist. H is called **grounded** if*

$$H(a_1, y) = H(x, a_2) = 0$$

for all $(x, y) \in S_1 \times S_2$.

Again, a_1 and a_2 can be $-\infty$.

If H is 2-increasing we have, from definition 2.1,

$$H(x_2, y_2) - H(x_1, y_2) \geq H(x_2, y_1) - H(x_1, y_1) \tag{2.1}$$

and

$$H(x_2, y_2) - H(x_2, y_1) \geq H(x_1, y_2) - H(x_1, y_1) \tag{2.2}$$

for every $[x_1, x_2] \times [y_1, y_2] \subset S_1 \times S_2$. By setting $x_1 = a_1$ in (2.1) and $y_1 = a_2$ in (2.2) we obtain the following lemma.

Lemma 2.4 *Any grounded, 2-increasing function $H : S_1 \times S_2 \to \mathcal{R}$ is nondecreasing in both arguments, that is, for all $x_1 \leq x_2$ in S_1 and $y_1 \leq y_2$ in S_2*

$$\begin{aligned} H(\,u\,, y_2) &\geq H(\,u\,, y_1), \\ H(x_2,\, v\,) &\geq H(x_1,\, v\,), \end{aligned}$$

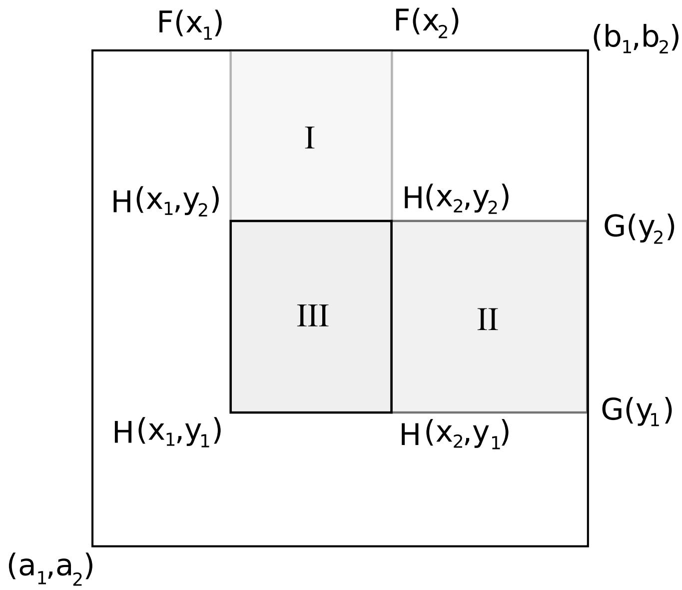


Figure 2.1: *Schematic proof of lemma 2.5. Apply 2-increasingness to rectangles I–III and combine the resulting inequalities. For the absolute value bars, use that H is nondecreasing in both arguments (lemma 2.4).*

for all u and all v in $\overline{\mathcal{R}}$.

From lemma 2.4 it follows that (2.1) and (2.2) also hold in absolute value:

$$
\begin{array}{rll}
|F(x_2) - F(x_1)| \geq |H(x_2, y_2) - H(x_1, y_2)| & \{ \text{ Ineq. (2.1) applied to rectangle I, Fig. 2.1 } \} \\
|G(y_2) - G(y_1)| \geq |H(x_2, y_2) - H(x_2, y_1)| & \{ \text{ Ineq. (2.2) applied to rectangle II, Fig. 2.1 } \} \\
\hline & + \\
|F(x_2) - F(x_1)| + |G(y_2) - G(y_1)| \;\geq\; |\,2H(x_2, y_2) - H(x_1, y_2) - H(x_2, y_1)\,| & \{ \text{ Triangle ineq. } \}
\end{array}
$$

Applying the definition of a 2-increasing function to rectangle III, Figure 2.1, yields

$$2H(x_2, y_2) - H(x_1, y_2) - H(x_2, y_1) \geq H(x_2, y_2) - H(x_1, y_1) \geq 0\,.$$

We thus have the next lemma.

Lemma 2.5 *For any grounded, 2-increasing function $H : S_1 \times S_2 \to \mathcal{R}$,*

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)|$$

for every $[x_1, x_2] \times [y_1, y_2] \subset S_1 \times S_2$.

Definition 2.6 *A grounded, 2-increasing function $C' : S_1 \times S_2 \to \mathcal{R}$ where S_1 and S_2 are subsets of $[0, 1]$ containing 0 and 1, is called a (two dimensional)* ***subcopula*** *if for all $(u, v) \in S_1 \times S_2$*

$$C'(u, 1) = u,$$
$$C'(1, v) = v.$$

Definition 2.7 *A (two dimensional)* ***copula*** *is a subcopula whose domain is* $[0,1]^2$.

Remark 2.8 *Note that reformulating lemma 2.5 in terms of subcopulas immediately leads to the Lipschitz condition*

$$|C'(u_2,v_2) - C'(u_1,v_1)| \leq |u_2 - u_1| + |v_2 - v_1|, \quad (u_1,v_1),(u_2,v_2) \in S_1 \times S_2,$$

which guarantees continuity of (sub)copulas.

Definition 2.9 *If* C *is differentiable on* $S_1 \times S_2$*, then the* ***density*** *associated with* C *is*

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v}.$$

2.1 Sklar's theorem

The theorem under consideration in this section, due to Sklar in 1959, is the very reason why copulas are popular for modeling purposes. It says that every joint distribution with continuous margins can be uniquely written as a copula function of its marginal distributions. This provides a way to separate the study of joint distributions into the marginal distributions and their joining copula.

Following Nelsen [2], we state Sklar's theorem for subcopulas first, the proof of which is short. The corresponding result for copulas follows from a straightforward, but elaborate, extension that will be omitted.

Definition 2.10 *Given a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *— where* Ω *is the sample space,* $\mathbb{P}$ *a measure such that* $\mathbb{P}(\Omega) = 1$ *and* $\mathcal{F} \subset 2^{\Omega}$ *a sigma-algebra — a* ***random variable*** *is defined to be a mapping*

$$X : \Omega \rightarrow \mathcal{R}$$

such that X *is* $\mathcal{F}$*-measurable.*

Definition 2.11 *Let* X *be a random variable. The* ***cumulative distribution function*** *(CDF) of* X *is*

$$F : \mathcal{R} \rightarrow [0,1], \quad F(x) := \mathbb{P}[X \leq x].$$

This will be denoted "$X \sim F$*".*

Definition 2.12 *If the derivative of the CDF of* X *exists, it is called the* ***probability density function*** *(pdf) of* X.

Definition 2.13 *Let* X *and* Y *be random variables. The* ***joint distribution function*** *of* X *and* Y *is*

$$H(x,y) := \mathbb{P}[X \leq x, Y \leq y].$$

The ***margins*** *of* H *are* $F(x) := \lim_{y\to\infty} H(x,y)$ *and* $G(y) := \lim_{x\to\infty} H(x,y)$.

Definition 2.14 *A random variable is said to be **continuous** if its CDF is continuous.*

Lemma 2.15 *Let H be a joint distribution function with margins F and G. Then there exists a unique subcopula C' such that*

$$Dom\, C' = Ran\, F \times Ran\, G$$

and

$$H(x,y) = C'(F(x), G(y)) \tag{2.3}$$

for all $(x,y) \in \overline{\mathcal{R}}$.

Proof

Unicity:

For C' to be unique, every $(u,v) \in \text{Ran}F \times \text{Ran}G$ should have only one possible image $C'(u,v)$ that is consistent with (2.3). Suppose to the contrary that $C'_1(u,v) \neq C'_2(u,v)$ are both consistent with (2.3), i.e. there exist $(x_1,y_1),(x_2,y_2) \in \overline{\mathcal{R}}^2$ such that

$$H(x_1,y_1) = C'_1(F(x_1), G(y_1)) = C'_1(u,v),$$

$$H(x_2,y_2) = C'_2(F(x_2), G(y_2)) = C'_2(u,v).$$

Thus, it must hold that $u = F(x_1) = F(x_2)$ and $v = G(y_1) = G(y_2)$. Being a joint CDF, H satisfies the requirements of lemma 2.5 and this yields

$$|H(x_2,y_2) - H(x_1,y_1)| \leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)| = 0,$$

so C'_1 and C'_2 agree on (u,v).

Existence:

Now define C' to be the (unique) function mapping the pairs $(F(x), G(y))$ to $H(x,y)$, for $(x,y) \in \overline{\mathcal{R}}^2$. It remains to show that C' is a 2-subcopula.

Groundedness:

$$C'(0, G(y)) = C'(F(-\infty), G(y)) = H(-\infty, y) = 0$$
$$C'(F(x), 0) = C'(F(x), G(-\infty)) = H(x, -\infty) = 0$$

2-increasingness:

Let $u_1 \leq u_2$ be in Ran F and $v_1 \leq v_2$ in Ran G. As CDFs are nondecreasing, there exist unique $x_1 \leq x_2, y_1 \leq y_2$ with $F(x_1) = u_1$, $F(x_2) = u_2$, $G(y_1) = v_1$ and $G(y_2) = v_2$.

$$\begin{aligned} & C'(u_2,v_2) - C'(u_1,v_2) - C'(u_2,v_1) + C'(u_1,v_1) \\ = \; & C'(F(x_2),G(y_2)) - C'(F(x_1),G(y_2)) - C'(F(x_2),G(y_1)) + C'(F(x_1),G(y_1)) \\ = \; & H(u_2,v_2) - H(u_1,v_2) - H(u_2,v_1) + H(u_1,v_1) \geq 0 \end{aligned}$$

The last inequality follows from the sigma-additivity of $\mathbb{P}$.

Margins are the identity mapping:

$$C'(1, G(y)) = C'(F(\infty), G(y)) = H(\infty, y) = G(y)$$

$$C'(F(x), 1) = C'(F(x), G(\infty)) = H(x, \infty) = F(x) \qquad \square$$

Remark 2.16 *The converse of lemma 2.15 also holds: every H defined by (2.3) is a joint distribution. This follows from the properties of a subcopula.*

Theorem 2.17 (Sklar's theorem) *Let H be a joint distribution function with margins F and G. Then there exists a 2-copula C such that for all $(x, y) \in \overline{\mathcal{R}}^2$*

$$H(x, y) = C(F(x), G(y)). \tag{2.4}$$

If F and G are continuous then C is unique.

Conversely, if F and G are distribution functions and C is a copula, then H defined by (2.4) is a joint distribution function with margins F and G.

Proof Lemma 2.15 provides us with a unique subcopula C' satisfying (2.4). If F and G are continuous, then $\text{Ran}F \times \text{Ran}G = I^2$ so $C := C'$ is a copula. If not, it can be shown (see [2]) that C' can be extended to a copula C.

The converse is a restatement of remark 2.16 for copulas. □

Now that the connection between random variables and copulas is established via Sklar's theorem, let us have a look at some implications.

Theorem 2.18 (C invariant under increasing transformations of X and Y) *Let $X \sim F$ and $Y \sim G$ be random variables with copula C. If α, β are increasing functions on Ran$X$ and Ran$Y$, then $\alpha \circ X \sim F \circ \alpha^{-1} := F_\alpha$ and $\beta \circ Y \sim G \circ \beta^{-1} := G_\beta$ have copula $C_{\alpha\beta} = C$.*

Proof

$$\begin{aligned} C_{\alpha\beta}(F_\alpha(x), G_\beta(y)) &= \mathbb{P}[\alpha \circ X \leq x, \beta \circ Y \leq y] = \mathbb{P}[X < \alpha^{-1}(x), Y < \beta^{-1}(y)] \\ &= C(F \circ \alpha^{-1}(x), G \circ \beta^{-1}(y)) = C(\mathbb{P}[X < \alpha^{-1}(x)], \mathbb{P}[Y < \beta^{-1}(y)]) \\ &= C(\mathbb{P}[\alpha \circ X < x], \mathbb{P}[\beta \circ Y < y]) = C(F_\alpha(x), G_\beta(y)) \quad \square \end{aligned}$$

Let $X \sim F$ and $Y \sim G$ be continuous random variables with joint distribution H. X and Y are independent iff. $H(x, y) = F(x)G(y)$. In terms of copulas this reads

Remark 2.19 *The continuous random variables X and Y are independent if and only if their copula is $C^\perp(u, v) = uv$.*

$C^\perp$ is called the **product copula**.

2.2 Fréchet-Hoeffding bounds

In this section we will show the existence of a maximal and a minimal bivariate copula, usually referred to as the Fréchet-Hoeffing bounds. All other copulas take values in between these bounds on each point of their domain, the unit square. The Fréchet upper bound corresponds to perfect positive dependence and the lower bound to perfect negative dependence.

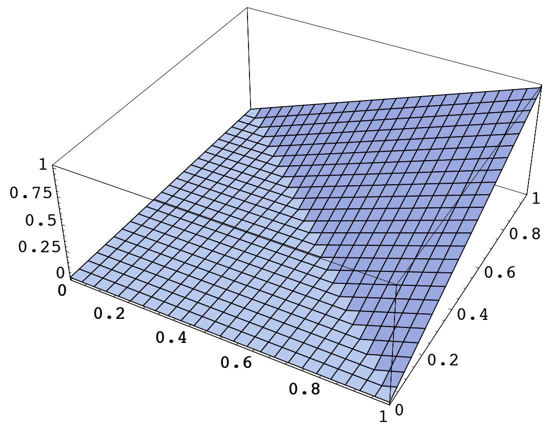


Figure 2.2: *Fréchet-Hoeffding lower bound*

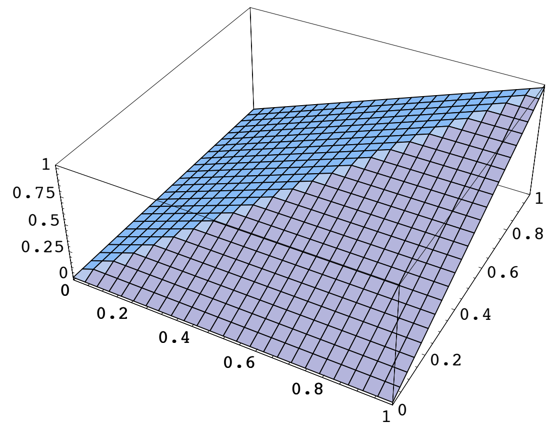


Figure 2.3: *Fréchet-Hoeffding upper bound*

Theorem 2.20 *For any subcopula C' with domain $S_1 \times S_2$*

$$C^-(u,v) := \max(u+v-1,0) \leq C'(u,v) \leq \min(u,v) =: C^+(u,v),$$

for every $(u,v) \in S_1 \times S_2$. C^+ and C^- are called the ***Fréchet-Hoeffding*** *upper and lower bounds respectively.*

Proof From lemma 2.4 we have $C'(u,v) \leq C'(u,1) = u$ and $C'(u,v) \leq C'(1,v) = v$, thus the upper bound.

$V_H([u,1] \times [v,1]) \geq 0$ gives $C'(u,v) \geq u+v-1$ and $V_H([0,u] \times [0,v]) \geq 0$ leads to $C'(u,v) \geq 0$. Combining these two gives the lower bound. $\square$

Plots of C^+ and C^- are provided in Figures 2.2 and 2.3. The remaining part of this section is devoted to the question under what condition these bounds are attained.

Definition 2.21 *A set $S \subset S_1 \times S_2 \subset \overline{\mathcal{R}}^2$ is called* ***nondecreasing*** *if for every $(x_1,y_1),(x_2,y_2) \in S$ it holds that $x_1 < x_2 \Rightarrow y_1 \leq y_2$. S is called* ***nonincreasing*** *if $x_1 > x_2 \Rightarrow y_1 \leq y_2$.*

An example of a nondecreasing set can be found in Figure 2.4.

Definition 2.22 *The* ***support*** *of a distribution function H is the complement of the union of all open subsets of $\mathcal{R}^2$ with H-volume zero.*

Remark 2.23 *Why not define the support of a distribution as the set where the joint density function is non-zero?*

1. *The joint density does not necessarily exist.*
2. *The joint density can be non-zero in isolated points. These isolated points are not included in definition 2.22.*

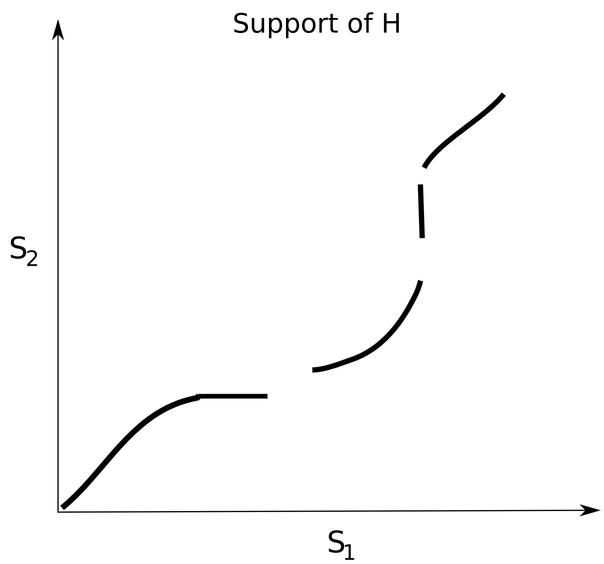


Figure 2.4: *Example of a nondecreasing set.*

Let X and Y be random variables with joint distribution H and continuous margins $F : S_1 \rightarrow \mathcal{R}$ and $G : S_1 \rightarrow \mathcal{R}$. Fix $(x, y) \in \overline{\mathcal{R}}^2$. Suppose H is equal to the Fréchet upper bound, then either $H(x, y) = F(x)$ or $H(x, y) = G(y)$. On the other hand we have

$$\begin{aligned} F(x) &= H(x,y) + \mathbb{P}[X \leq x, Y > y], \\ G(y) &= H(x,y) + \mathbb{P}[X > x, Y \leq y]. \end{aligned}$$

It follows that either $\mathbb{P}[X \leq x, Y > y]$ or $\mathbb{P}[X > x, Y \leq y]$ is zero. As suggested by Figure 2.5 this can only be true if the support of H is a nondecreasing set.

This intuition is confirmed by the next theorem, a proof of which can be found in Nelsen [2].

Theorem 2.24 *Let X and Y be random variables with joint distribution function H.*

H is equal to the upper Fréchet-Hoeffding bound if and only if the support of H is a nondecreasing subset of $\overline{\mathcal{R}}^2$.

H is equal to the lower Fréchet-Hoeffding bound if and only if the support of H is a nonincreasing subset of $\overline{\mathcal{R}}^2$.

Remark 2.25 *If X and Y are* ***continuous*** *random variables, then the support of H cannot have horizontal or vertical segments. Indeed, suppose the support of H would have a horizontal line segment, then a relation of the form $0 < \mathbb{P}[a \leq X \leq b] = \mathbb{P}[Y = c]$ would hold, implying that the CDF of Y had a jump at c.*

Thus, in case of continuous X and Y, theorem 2.24 implies the support of H to be an almost surely increasing (decreasing) set if and only if H is equal to the upper (lower) Fréchet-Hoeffding bound.

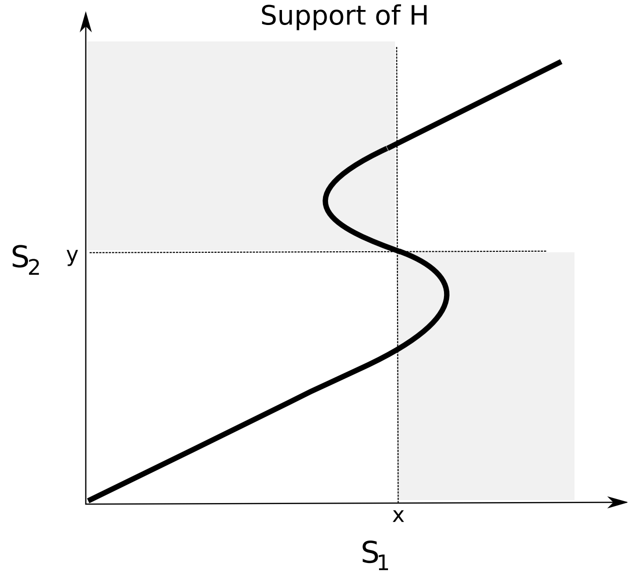


Figure 2.5: *In case of non-perfect positive dependence, the shaded area always contains points with nonzero probability.*

Remark 2.26 *The support of H being an almost surely (in)(de)creasing set means that if you observe X, there is only one Y that can be observed simultaneously, and vice versa. Intuitively, this is exactly the notion of 'perfect dependence'.*

2.3 Survival copula

Every copula has a survival copula associated with it which is a mirrored version of the original copula. Particularly useful is the fact that its upper and lower tail properties are interchanged.

Let $X \sim F$ and $Y \sim G$ be random variables with copula C. The joint survival function for the vector $(F(X), G(Y))$ of uniform random variables represents, when evaluated in (u, v), the joint probability that $(F(X), G(Y))$ be greater (component-wise) than (u, v). Due to Sklar's theorem this joint survival function is a copula. It is called the **survival copula** of C.

Lemma 2.27 *The survival copula $\overline{C}$ associated with the copula C satisfies*

$$\overline{C}(u, v) = 1 - u - v + C(u, v).$$

Proof

$$\begin{aligned} \overline{C}(u, v) &=: \mathbb{P}[F(X) > u, G(Y) > v] \\ &= 1 - \mathbb{P}[F(X) < u] - \mathbb{P}[G(Y) < v] + \mathbb{P}[F(X) < u, G(Y) < v] \\ &= 1 - u - v + C(u, v)\,. \end{aligned}$$

3 Multivariate Copulas

The notion of copulas, introduced in section 2, will now be generalized to dimensions $n \geq 2$. This we will need to price derivatives on more than two underlyings.

The majority of the results of the previous section have equivalents in the multivariate case, an exception being the generalized Fréchet-Hoeffding lower bound, which is not a copula for $n \geq 3$.

Definition 3.1 *Let H be an $S_1 \times S_2 \times \ldots \times S_n \to \mathcal{R}$ function, where the non-empty sets $S_i \subset \overline{\mathcal{R}}$ have minimum a_i and maximum b_i, $1 \leq i \leq n$. H is called* ***grounded*** *if for every u in the domain of H that has at least one index k such that $u_k = a_k$:*

$$H(u) = H(u_1, \ldots, u_{k-1}, a_k, u_{k+1}, \ldots, u_n) = 0.$$

Definition 3.2 *Let $x, y \in \overline{\mathcal{R}}^n$ such that $x \leq y$ holds component-wise. Define the* ***n-box*** *$[x, y]$ by*

$$[x, y] := [x_1, y_1] \times [x_2, y_2] \times \ldots \times [x_n, y_n].$$

The set of vertices $\text{ver}([x, y])$ of $[x, y]$ consists of the 2^n points w that have $w_i = x_i$ or $w_i = y_i$ for $1 \leq i \leq n$. The product

$$\text{sgn}(w) := \prod_{i=1}^{2^n} \text{sgn}(2w_i - x_i - y_i)$$

equals 0 if $x_i = y_i$ for some $1 \leq i \leq n$. If $\text{sgn}(w)$ is non-zero, it equals $+1$ if $w - x$ has an even number of zero components and -1 if $w - x$ has an odd number of zero components.

Using this inclusion-exclusion idea, we can now define n-increasingness:

Definition 3.3 *The function $H : S_1 \times \ldots \times S_n \to \mathcal{R}$ is said to be* ***n-increasing*** *if the H-volume of every n-box $[x, y]$ with $\text{ver}([\text{x,y}]) \in S_1 \times \ldots \times S_n$ is nonnegative:*

$$\sum_{w \in \text{ver}([x,y])} \text{sgn(w)} H(w) \geq 0 \tag{3.1}$$

Definition 3.4 *The* ***k-dimensional margins*** *of $H : S_1 \times \ldots \times S_n \to \mathcal{R}$ are the functions $F_{i_1 i_2 \ldots i_k} : S_{i_1} \times \ldots \times S_{i_k} \to \mathcal{R}$ defined by*

$$F_{i_1 i_2 \ldots i_k}(u_{i_1}, \ldots, u_{i_k}) = H(b_1, b_2, \ldots, u_{i_1}, \ldots, u_{i_2}, \ldots, u_{i_k}, \ldots, b_n).$$

Definition 3.5 *A grounded, n-increasing function $C' : S_1 \times \ldots \times S_n \to \mathcal{R}$ is an* ***n-dimensional subcopula*** *if each S_i contains at least 0 and 1 and all one-dimensional margins are the identity function.*

Definition 3.6 *An n-dimensional subcopula for which $S_1 \times \ldots \times S_n = I^n$ is an* ***n-dimensional copula****.*

3.1 Sklar's theorem

Theorem 3.7 (Sklar's theorem, multivariate case) *Let H be an n-dimensional distribution function with margins $F_1, \ldots, F_n$. Then there exists an n-copula C such that for all $u \in \overline{\mathcal{R}}^n$*

$$H(u_1, \ldots, u_n) = C(F(u_1), \ldots, F(u_n)). \tag{3.2}$$

If $F_1, \ldots, F_n$ are continuous, then C is unique.

Conversely, if $F_1, \ldots, F_n$ are distribution functions and C is a copula, then H defined by (3.2). is a joint distribution function with margins $F_1, \ldots, F_n$.

3.2 Fréchet-Hoeffding bounds

Theorem 3.8 *For every copula C and any $u \in I^n$*

$$C^-(u) := \max(u_1 + u_2 + \ldots + u_n - n + 1, 0) \leq C(u) \leq \min(u_1, u_2, \ldots, u_n) := C^+(u).$$

In the multidimensional case, the upper bound is still a copula, but the lower bound is not.

The following example, due to Schweizer and Sklar [3], shows that C^- does not satisfy equation (3.1). Consider the n-box $\left[\frac{1}{2}, 1\right] \times \ldots \times \left[\frac{1}{2}, 1\right]$. For 2-increasingness, in particular, the H-volume of this n-box has to be nonnegative. This is not the case for $n > 2$:

$$\begin{aligned}
&\underbrace{\max\left\{1 + \ldots + 1 - n + 1, 0\right\}}_{=n-n+1=1} - n \underbrace{\max\left\{\frac{1}{2} + 1 + \ldots + 1 - n + 1, 0\right\}}_{=\frac{1}{2}+(n-1)-n+1=\frac{1}{2}} \\
&+ \binom{n}{2} \underbrace{\max\left\{\frac{1}{2} + \frac{1}{2} + 1 + \ldots + 1 - n + 1, 0\right\}}_{=0} + \ \ldots\ldots \ \pm \underbrace{\max\left\{\frac{1}{2} + \ldots + \frac{1}{2} - n + 1, 0\right\}}_{=0} \\
&= \ 1 - \frac{n}{2}.
\end{aligned}$$

On the other hand, for every $u \in I^n$, $n \geq 3$, there exists a copula C such that $C(u) = C^-(u)$ (see Nelsen [2]). This shows that a sharper lower bound does not exist.

3.3 Survival copula

Analogous to the bivariate case (Section 2.3) one can define a multivariate survival copula. Let $X_i \sim F_i$, $1 \leq i \leq n$, be random variables with copula C. The joint survival function for the vector $(\,F_1(X_1), \ldots, F_n(X_n)\,)$ of uniform random variables represents, when evaluated in $(u_1, \ldots, u_n)$, the joint probability that $(\,F_1(X_1), \ldots, F_n(X_n)\,)$ be greater (component-wise) than $(u_1, \ldots, u_n)$. Due to the multivariate version of Sklar's theorem, this joint survival function is a copula. It is called the survival copula of C.

4 Dependence

The dependence structure between random variables is completely described by their joint distribution function. ‘Benchmarks’ like linear correlation only capture certain parts of this dependence structure. Apart from linear correlation, there exist several other **measures of association**. These, and their relation to copulas, are the subject of this section.

Scarsini [4] describes measures of association as follows:

“*Dependence is a matter of association between X and Y along any measurable function, i.e. the more X and Y tend to cluster around the graph of a function, either $y = f(x)$ or $x = g(y)$, the more they are dependent.*”

There exists some freedom in how to define the ‘extent to which X and Y cluster around the graph of a function’. The choice of this function is exactly the point where the most important measures of association differ.

Section 4.1 explains the concept of linear correlation. It measures how well two random variables cluster around a **linear** function. A major shortcoming is that linear correlation is not invariant under non-linear monotonic transformations of the random variables.

The concordance and dependence measures (e.g. Kendall’s tau, Spearman’s rho) introduced in sections 4.2 and 4.3 reflect the degree to which random variables cluster around a **monotone** function. This is a consequence of these measures being defined such as only to depend on the copula — see definition 4.5(6) — and copulas are invariant under monotone transformations of the random variables.

Finally, in section 4.4 dependence will be studied in case the involved random variables simultaneously take extreme values.

From now on the random variables X and Y are assumed to be continuous.

4.1 Linear correlation

Definition 4.1 *For non-degenerate, square integrable random variables X and Y the* ***linear correlation coefficient*** *ρ is*

$$\rho = \frac{Cov[X,Y]}{(Var[X]\,Var[Y])^{\frac{1}{2}}}$$

Correlation can be interpreted as the degree to which a linear relation succeeds to describe the dependency between random variables. If two random variables are linearly dependent, then $\rho = 1$ or $\rho = -1$.

Example 4.2 *Let X be a uniformly distributed random variable on the interval $(0, 1)$ and set $Y = X^n$, $n \geq 1$. X and Y thus are perfectly positive dependent.*

The n-th moment of X is

$$\mathbb{E}\left[X^n\right] = \int_0^1 x^n\,dx = \frac{1}{1+n}. \tag{4.1}$$

The linear correlation between X and Y is

$$\begin{aligned}\rho &= \frac{\mathbb{E}[XY]-\mathbb{E}[X]\,\mathbb{E}[Y]}{(\mathbb{E}[X^2]-\mathbb{E}[X]^2)^{\frac{1}{2}}(\mathbb{E}[Y^2]-\mathbb{E}[Y]^2)^{\frac{1}{2}}} \\ &= \frac{\mathbb{E}\left[X^{n+1}\right]-\mathbb{E}[X]\,\mathbb{E}[X^n]}{(\mathbb{E}[X^2]-\mathbb{E}[X]^2)^{\frac{1}{2}}(\mathbb{E}[X^{2n}]-\mathbb{E}[X^n]^2)^{\frac{1}{2}}} \\ &\overset{(4.1)}{=} \frac{\sqrt{3+6n}}{2+n}.\end{aligned}$$

For $n = 1$ the correlation coefficient equals 1, for $n > 1$ it is less than 1.

Corollary 4.3 *From the above example we conclude:*

(i). The linear correlation coefficient is not invariant under increasing, non-linear transforms.

(ii). Random variables whose joint distribution has nondecreasing or nonincreasing support can have correlation coefficient different from 1 or −1.

4.2 Measures of concordance

Consider the following definition of concordance and discordance. The first part applies to observations from a pair of random variables, the second part to copula functions.

Definition 4.4

(i). Two observations (x_1, y_1) and (x_2, y_2) are ***concordant*** *if $x_1 < x_2$ and $y_1 < y_2$ or if $x_1 > x_2$ and $y_1 > y_2$. An equivalent characterisation is $(x_1 - x_2)(y_1 - y_2) > 0$. The observations (x_1, y_1) and (x_2, y_2) are said to be* ***discordant*** *if $(x_1 - x_2)(y_1 - y_2) < 0$.*

(ii). If C_1 and C_2 are copulas, we say that C_1 is ***less concordant*** *than C_2 (or C_2 is* ***more concordant*** *than C_1) and write $C_1 \prec C_2$ ($C_2 \succ C_1$) if*

$$C_1(u) \leq C_2(u) \quad \text{and} \quad \overline{C}_1(u) \leq \overline{C}_2(u) \quad \text{for all } u \in I^m. \tag{4.2}$$

In the remaining part of this section we will only consider bivariate copulas. Part (ii) of definition 4.4 is then equivalent to $C_1(u, v) \leq C_2(u, v)$ for all $u \in I^2$, see lemma 2.27.

Definition 4.5 *A measure of association $\kappa_C = \kappa_{X,Y}$ is called a* ***measure of concordance*** *if*

1. *$\kappa_{X,Y}$ is defined for every pair X, Y of random variables,*

2. *$-1 \leq \kappa_{X,Y} \leq 1$, $\kappa_{X,X} = 1$, $\kappa_{-X,X} = -1$,*

3. *$\kappa_{X,Y} = \kappa_{Y,X}$,*

4. *if X and Y are independent then $\kappa_{X,Y} = \kappa_{C^\perp} = 0$,*

5. $\kappa_{-X,Y} = \kappa_{X,-Y} = -\kappa_{X,Y}$,

6. *if C_1 and C_2 are copulas such that $C_1 \prec C_2$ then $\kappa_{C_1} \leq \kappa_{C_2}$,*

7. *if $\{(X_n, Y_n)\}$ is a sequence of continuous random variables with copulas C_n and if $\{C_n\}$ converges pointwise to C, then $\lim_{n\to\infty} \kappa_{X_n,Y_n} = \kappa_C$.*

What is the connection between definition 4.4 and 4.5?

It is natural to think of a concordance measure as being defined by the copula only. Indeed, by applying axiom (6) twice it follows that $C_1 = C_2$ implies $\kappa_{C_1} = \kappa_{C_2}$. If the random variables X and Y have copula C and the transformations α and β are both strictly increasing, then $C_{X,Y} = C_{\alpha(X),\beta(Y)}$ by theorem 2.18 and consequently $\kappa_{X,Y} = \kappa_{\alpha(X),\beta(Y)}$. Via axiom (5) a similar result for strictly decreasing transformations can be established. **Measures of concordance thus are invariant under strictly monotone transformations of the random variables.**

If $Y = \alpha(X)$ and α is stictly increasing (decreasing), it follows from $C_{X,\alpha(X)} = C_{X,X}$ and axiom (2) that $\kappa_{X,Y} = 1 \ (-1)$. In other words: **a measure of concordance assumes its maximal (minimal) value if the support of the joint distribution function of X and Y contains only concordant (discordant) pairs.** This explains how definitions 4.4 and 4.5 are related.

Summarizing,

Lemma 4.6

(i). Measures of concordance are invariant under strictly monotone transformations of the random variables.

(ii). A measure of concordance assumes its maximal (minimal) value if the support of the joint distribution function of X and Y contains only concordant (discordant) pairs.

Note that these properties are in conflict with the conclusions in corollary 4.3 for the linear correlation coefficient. Linear correlation thus is not a measure of concordance.

In the remaining part of this section, two concordance measures will be described: Kendall's tau and Spearman's rho.

4.2.1 Kendall's tau

Let Q be the difference between the probability of concordance and discordance of two independent random vectors (X_1, Y_1) and (X_2, Y_2):

$$Q = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0]. \tag{4.3}$$

In case (X_1, Y_1) and (X_2, Y_2) are independent and identically distributed (i.i.d.) random vectors, the quantity Q is called **Kendall's tau** τ.

Given a sample $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ of n observations from a random vector (X, Y), an unbiased estimator for τ is

$$t := \frac{c - d}{c + d},$$

where c is the number of concordant pairs and d the number of discordant pairs in the sample.

Nelsen [2] shows that if (X_1, Y_1) and (X_2, Y_2) are independent random vectors with (possibly different) distributions H_1 and H_2, but with common margins F, G and copulas C_1, C_2

$$Q = 4 \iint_{I^2} C_2(u, v)\, dC_1(u, v) - 1. \tag{4.4}$$

It follows that the probability of concordance between two bivariate distributions (with common margins) minus the probability of discordance only depends on the copulas of each of the bivariate distributions.

Note that if $C_1 = C_2 := C$, then, since we already assumed common margins, the distributions H_1 and H_2 are equal which means that (X_1, Y_1) and (X_2, Y_2) are identically distributed. In that case, (4.4) gives Kendall's tau for the i.i.d. random vectors (X_1, Y_1), (X_2, Y_2) with copula C.

Furthermore it can be shown that

$$\tau = 1 - 4 \iint_{I^2} \frac{\partial C(u, v)}{\partial u} \frac{\partial C(u, v)}{\partial v}\, du\, dv. \tag{4.5}$$

In the particular case that C is absolutely continuous, the above relation can be deduced via integration by parts.

As an example of the use of (4.5), consider

Lemma 4.7 $\tau_C = \tau_{\overline{C}}$.

Proof

$$\begin{aligned} \tau_{\overline{C}} &= 1 - 4 \iint_{I^2} \frac{\partial \overline{C}}{\partial u} \frac{\partial \overline{C}}{\partial v}\, du\, dv \\ &= 1 - 4 \iint_{I^2} [1 - \frac{\partial C}{\partial u}][1 - \frac{\partial C}{\partial v}]\, du\, dv \\ &= \tau_C - 4 \iint_{I^2} [1 - \frac{\partial C}{\partial u} - \frac{\partial C}{\partial v}]\, du\, dv. \end{aligned} \tag{4.6}$$

The second term of the integrand of (4.6) reduces to

$$\iint_{I^2} \frac{\partial C}{\partial u}\, du\, dv = \int_0^1 C(1, v) - C(0, v)\, dv = \int_0^1 C(1, v)\, dv = \int_0^1 v\, dv = \frac{1}{2}.$$

Similarly,

$$\iint_{I^2} \frac{\partial C}{\partial v}\, du\, dv = \frac{1}{2}.$$

Substituting in (4.6) yields the lemma. $\square$

Scarsini [4] proves that axioms (1)–(7) of definition 4.5 are satisfied by Kendall's tau.

The next lemma holds for Kendall's tau, but not for concordance measures in general.

Lemma 4.8 *Let H be a joint distribution with copula C.*

$$\begin{aligned} C = C^+ \quad & \textit{iff.} \quad \tau = 1, \\ C = C^- \quad & \textit{iff.} \quad \tau = -1. \end{aligned}$$

Proof We will prove the first statement, $C = C^+$ iff. $\tau = 1$, via the following steps:

(i) $\tau = 1 \Rightarrow H$ has nondecreasing support

(ii) H has nondecreasing support $\Rightarrow H = C^+$

(iii) $H = C^+ \Rightarrow \tau = 1$

Step (ii) is immediate from theorem 2.24. Step (iii) follows from substitution of C^+ in formula (4.4) and straightforward calculation. This step in fact also follows from axiom (6) in definition 4.5.

It remains to show that $\tau = 1$ implies H to have a nondecreasing support. Suppose $\tau = 1$ and H does not have a nondecreasing support so that there exists at least one discordant pair $(x_1, y_1), (x_2, y_2)$ in H. Define $\delta := \frac{1}{2}|\min\{x_1 - x_2, y_1 - y_2\}|$ and

$$\begin{aligned} B_1 &:= \mathcal{B}_\delta(x_1, y_1), \\ B_2 &:= \mathcal{B}_\delta(x_2, y_2), \end{aligned}$$

where $\mathcal{B}_r(x, y)$ denotes an open 2-ball with radius r and centre (x, y). From the definition of the support of a CDF it follows that both $\mathbb{P}[B_1] > 0$ and $\mathbb{P}[B_2] > 0$. Because $\tau = 1$ we have from equation (4.3) that

$$\begin{aligned} 0 &= \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0] \\ &= \iint_{\text{supp}H} \mathbb{P}[(X_2 - X_1)(Y_2 - Y_1) < 0 | (X_1, Y_1) = (u, v)]\, dH(u, v) \\ &= \iint_{\text{supp}H} \Big\{ \mathbb{P}[X_2 > u, Y_2 < v] + \mathbb{P}[X_2 < u, Y_2 > v] \Big\}\, dH(u, v) \\ &\geq \iint_{B_1 \cup \text{supp}H} \Big\{ \mathbb{P}[X_2 > u, Y_2 < v] + \mathbb{P}[X_2 < u, Y_2 > v] \Big\}\, dH(u, v) \\ &\geq \iint_{B_1 \cup \text{supp}H} \mathbb{P}[B_2]\, dH(u, v) \quad = \quad \mathbb{P}[B_2] \iint_{B_1 \cup \text{supp}H} dH(u, v) \quad = \quad \mathbb{P}[B_1]\mathbb{P}[B_2]. \end{aligned}$$

This is a contradiction. □

4.2.2 Spearman's rho

Let (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3) be i.i.d. random vectors with common joint distribution H, margins F, G and copula C. Spearman's rho is defined to be proportional to the probability of concordance minus the probability of discordance of the pairs $(\underbrace{X_1, Y_1}_{\text{Joint distr. } H})$ and $(\underbrace{X_2, Y_3}_{\text{Independent}})$:

$$\rho_S = 3\ (\ \mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) < 0]\ ).$$

Note that X_2 and Y_3, being independent, have copula $C^{\perp}$. By (4.4), three times the concordance difference between C and $C^{\perp}$ is

$$\begin{aligned} \rho_S &= 3\left(\ 4\iint_{I^2} C(u,v)dC^{\perp}(u,v) - 1\ \right) \\ &= 12\iint_{I^2} C(u,v)\,du\,dv - 3. \end{aligned} \tag{4.7}$$

Spearman's rho satisfies the axioms in definition 4.5 (see Nelsen [2]).

Let $X \sim F$ and $Y \sim G$ be random variables with copula C, then Spearman's rho is equivalent to the linear correlation between $F(X)$ and $G(Y)$. To see this, recall from probability theory that $F(X)$ and $G(Y)$ are uniformly distributed on the interval $(0,1)$, so $\mathbb{E}[F(X)] = \mathbb{E}[G(Y)] = 1/2$ and $\mathrm{Var}[F(X)] = \mathrm{Var}[G(Y)] = 1/12$. We thus have

$$\begin{aligned} \rho_S &\overset{(4.7)}{=} 12\,\mathbb{E}[F(X), G(Y)] - 3 \\ &= \frac{\mathbb{E}[F(X), G(Y)] - (1/2)^2}{1/12} \\ &= \frac{\mathbb{E}[F(X), G(Y)] - \mathbb{E}[F(X)]\,\mathbb{E}[G(Y)]}{(\mathrm{Var}[F(X)]\mathrm{Var}[G(Y)])^{\frac{1}{2}}} \\ &= \frac{\mathrm{Cov}[F(X), G(Y)]}{(\mathrm{Var}[F(X)]\mathrm{Var}[G(Y)])^{\frac{1}{2}}}. \end{aligned}$$

Cherubini et al. [5] state that for Spearman's rho a statement similar to lemma 4.8 holds: $C = C^{\pm}$ iff. $\rho_S = \pm 1$.

4.2.3 Gini's gamma

Whereas Spearman's rho measures the concordance difference between a copula C and independence, **Gini's gamma** γ_C measures the concordance difference between a copula C and monotone dependence, i.e. the copulas C^+ and C^- (Section 2.2),

$$\begin{aligned} \gamma_C &\overset{(4.4)}{=} \iint_{I^2} C(u,v)dC^-(u,v) \ + \ \iint_{I^2} C(u,v)dC^+(u,v) \\ &\overset{[2,\text{ Corollary 5.1.14}]}{=} 4\left[\int_0^1 C(u,1-u)du - \int_0^1 \Big(u - C(u,u)\Big)du\right]. \end{aligned}$$

Gini's gamma thus can be interpreted as the area between the secondary diagonal sections of C and C^-,

$$C(u,1-u) - C^-(u,1-u) = C(u,1-u) - \max(u + (1-u) - 1, 0) = C(u,1-u)\,,$$

minus the area between the diagonal sections of C^+ and C,

$$C^+(u,u) - C(u,u) = \min(u,u) - C(u,u) = u - C(u,u)\,.$$

4.3 Measures of dependence

Definition 4.9 *A measure of association $\delta_C = \delta_{X,Y}$ is called a* ***measure of dependence*** *if*

1. *$\delta_{X,Y}$ is defined for every pair X, Y of random variables,*
2. *$0 \leq \delta_{X,Y} \leq 1$*
3. *$\delta_{X,Y} = \delta_{Y,X}$,*
4. *$\delta_{X,Y} = 0$ iff. X and Y are independent,*
5. *$\delta_{X,Y} = 1$ iff. $Y = f(X)$ where f is a strictly monotone function,*
6. *if α and β are strictly monotone functions on Ran X and Ran Y respectively, then $\delta_{X,Y} = \delta_{\alpha(X),\beta(Y)}$,*
7. *if $\{(X_n, Y_n)\}$ is a sequence of continuous random variables with copulas C_n and if $\{C_n\}$ converges pointwise to C, then $\lim_{n\to\infty} \delta_{X_n,Y_n} = \delta_C$.*

The differences between dependence and concordance measures are:

(i). Concordance measures assume their maximal (minimal) values if the concerning random variables are perfectly positive (negative) dependent. Dependence measures assume their extreme values if **and only if** the random variables are perfectly dependent.

(ii). Concordance measures are zero in case of independence. Dependence measures are zero if **and only if** the random variables under consideration are independent.

(iii). The stronger properties of dependence measures over concordance measures go at the cost of a sign: dependence is a measure of association with respect to a monotone function — indifferently increasing or decreasing — whereas concordance accounts for the kind of monotonicity [4].

4.3.1 Schweizer and Wolff's sigma

Schweizer and Wolff's sigma for two random variables with copula C is given by

$$\sigma_C = 12 \iint_{I^2} \left| C(u,v) - uv \right| du\, dv.$$

Nelsen [2] shows that this association measure satisfies the properties of definition 4.9.

4.4 Tail dependence

This section examines dependence in the upper-right and lower-left quadrant of I^2.

Definition 4.10 *Given two random variables $X \sim F$ and $Y \sim G$ with copula C, define the coefficients of tail dependence*

$$\lambda_L := \lim_{u \downarrow 0} \mathbb{P}[F(X) < u | G(Y) < u] = \lim_{u \downarrow 0} \frac{C(u,u)}{u}, \tag{4.8}$$

$$\lambda_U := \lim_{u \uparrow 1} \mathbb{P}[F(X) > u | G(Y) > u] = \lim_{u \uparrow 1} \frac{1 - 2u + C(u,u)}{1-u}. \tag{4.9}$$

C is said to have ***lower (upper) tail dependence*** *iff. $\lambda_L \neq 0$ $(\lambda_U \neq 0)$.*

The coefficients of tail dependence express the probability of two random variables both taking extreme values.

Lemma 4.11 *Denote the lower (upper) coefficient of tail dependence of the survival copula $\overline{C}$ by $\overline{\lambda}_L$ $(\overline{\lambda}_U)$, then*

$$\lambda_L = \overline{\lambda}_U,$$
$$\lambda_U = \overline{\lambda}_L.$$

Proof

$$\lambda_L = \lim_{u \downarrow 0} \frac{C(u,u)}{u} = \lim_{v \uparrow 1} \frac{C(1-v, 1-v)}{1-v} = \lim_{v \uparrow 1} \frac{1 - 2v + \overline{C}(v,v)}{1-v} = \overline{\lambda}_U$$

$$\lambda_U = \lim_{u \uparrow 1} \frac{1 - 2u + C(u,u)}{1-u} = \lim_{u \downarrow 0} \frac{2v - 1 + C(1-v, 1-v)}{v} = \lim_{u \downarrow 0} \frac{\overline{C}(v,v)}{v} = \overline{\lambda}_L \qquad \square$$

Example 4.12 *As an example, consider the* ***Gumbel copula***

$$C_{Gumbel}(u,v) := \exp\{-[(-\log u)^{\frac{1}{\alpha}} + (-\log v)^{\frac{1}{\alpha}}]^{\alpha}\}, \quad \alpha \in [1, \infty)$$

with ***diagonal section***

$$\tilde{C}_{Gumbel}(u) := C_{Gumbel}(u,u) = u^{2^{\alpha}}.$$

$\tilde{C}$ is differentiable in both $u = 0$ and $u = 1$, this is a sufficient condition for the limits (4.8) and (4.9) to exist:

$$\begin{aligned} \lambda_L &= \frac{d\tilde{C}}{du}(0) \\ &= \left[\frac{d}{du} u^{2^{\alpha}}\right]_{u=0} = \left[2^{\alpha} u^{2^{\alpha}-1}\right]_{u=0} = 0, \end{aligned}$$

$$\begin{aligned} \lambda_U &= \overline{\lambda}_L = \frac{d}{du}\left[2\,u - 1 + \tilde{C}(1-u)\right]_{u=0} = 2 - \frac{d\tilde{C}}{du}(1) \\ &= 2 - \left[\frac{d}{du} u^{2^{\alpha}}\right]_{u=1} = 2 - \left[2^{\alpha} u^{2^{\alpha}-1}\right]_{u=1} = 2 - 2^{\alpha}. \end{aligned}$$

So the Gumbel copula has no lower tail dependency. It has upper tail dependency iff. $\alpha \neq 1$.

4.5 Multivariate dependence

Most of the concordance and dependence measures introduced in the previous sections have one or more multivariate generalizations.

Joe [6] obtains the following generalized version of Kendall's tau. Let $X = (X_1, \ldots, X_m)$ and $Y = (Y_1, \ldots, Y_m)$ be i.i.d. random vectors with copula C and define $D_j := X_j - Y_j$. Denote by $B_{k,m-k}$ the set of m-tuples in $\mathcal{R}^m$ with k positive and $m-k$ negative components. A generalized version of Kendall's tau is given by

$$\tau_C = \sum_{k=\lfloor \frac{m+1}{2} \rfloor}^{m} w_k \, \mathbb{P}(\,(D_1, \ldots, D_m) \in B_{k,m-k})$$

where the weights w_k, $\lfloor \frac{m+1}{2} \rfloor \leq k \leq m$, are such that

(i). $\tau_C = 1$ if $C = C^+$,

(ii). $\tau_C = 0$ if $C = C^\perp$,

(iii). $\tau_{C_1} < \tau_{C_2}$ whenever $C_1 \prec C_2$.

The implications of (i) and (ii) for the w_k's are straightforward:

(i). $w_m = 1$,

(ii). $\sum_{k=0}^{m} w_k \binom{m}{k} = 0 \quad (w_k := w_{m-k}$ for $k < \lfloor \frac{m+1}{2} \rfloor)$.

The implication of (iii) is more involved (see [6, p. 18]), though it is clear that at least $w_m \geq w_{m-1} \geq \ \ldots \ \geq w_{\lfloor \frac{m+1}{2} \rfloor}$ should hold.

For $m = 3$ the only weights satisfying (i)–(iii) are $w_3 = 1$ and $w_2 = -\frac{1}{3}$. The minimal value of τ for $m = 3$ thus is $-\frac{1}{3}$. For $m = 4$ there exists a one-dimensional family of generalizations of Kendall's tau.

In terms of copulas, Joe's generalization of Spearman's rho [6, pp. 22-24] for a m-multivariate distribution function having copula C reads

$$\omega_C = \left(\int \cdots \int_{I^m} C(u)\, du_1 \ldots du_m \; - \; 2^{-m} \right) \Big/ \left((m+1)^{-1} + 2^{-m} \right).$$

Properties (i) and (ii) are taken care of by the scaling and normalization constants and can be checked by substituting C^+ and $C^\perp$. The increasingness of ω with respect to $\prec$ is immediate from definition 4.4(ii).

There also exist multivariate measures of dependence. For instance, Nelsen [2] mentions the following generalization of Schweizer and Wolff's sigma:

$$\sigma_C = \frac{2^m(m+1)}{2^m - (m+1)} \int \cdots \int_{I^m} |C(u_1, \ldots, u_m) - u_1 \ldots u_m| \, du_1 \ldots du_m,$$

where C is an m-copula.

5 Parametric families of copulas

This section gives an overview of some types of parametric families of copulas. We are particularly interested in their coefficients of tail dependence.

The Fréchet family (section 5.1) arises by taking affine linear combinations of the product copula and the Fréchet-Hoeffding upper and lower bounds. Tail dependence is determined by the weights in the linear combination.

In section 5.2 copulas are introduced which stem from elliptical distributions. Because of their symmetric nature, upper and lower tail dependence coefficients are equal.

Any function satisfying certain properties (described in section 5.3) generates an Archimedean copula. These copulas can take a great variety of forms. Furthermore, they can have distinct upper and lower tail dependence coefficients. This makes them suitable candidates for modeling asset prices, since in market data either upper or lower tail dependence tends to be more profound.

Multivariate Archimedean copulas however are of limited use in practice as all bivariate margins are equal. Therefore in section 5.4 an extension of the class of Archimedean copulas will be discussed that allows for several distinct bivariate margins.

5.1 Fréchet family

Every affine linear combination of copulas is a new copula. This fact can be used for instance to construct the Fréchet family of copulas

$$\begin{aligned} C^F(u,v) &= pC^-(u,v) + (1-p-q)C^\perp(u,v) + qC^+(u,v) \\ &= p\max(u+v-1,0) + (1-p-q)uv + q\min(u,v) \end{aligned}$$

where $C^\perp(u,v) = uv$ is the product copula and $0 \le p, q, \le 1$, $p+q \le 1$.

The product copula models independence, whereas the Fréchet-Hoeffding upper and lower bounds 'add' positive and negative dependence respectively. This intuition is confirmed by Spearman's rho:

$$\begin{aligned} \rho_{S\,C^F} &= 12 \iint_{I^2} C^F(u,v)\,du\,dv - 3 \\ &= 12 \int_0^1 \int_{1-v}^1 p(u+v-1)\,du\,dv + 12\,(1-p-q) \iint_{I^2} uv\,du\,dv \\ &\quad + 12 \int_0^1 \int_0^u qv\,dv\,du + 12 \int_0^1 \int_u^1 qu\,dv\,du - 3 \\ &= q - p. \end{aligned}$$

Indeed, the weight p (of C^-) has negative sign and q (of C^+) has positive sign.

The Fréchet family has upper and lower tail dependence coefficient q.

5.2 Elliptical distributions

The n-dimensional random vector X is said to follow an elliptical distribution if $X - \mu$, for some $\mu \in \mathcal{R}^n$, has a characteristic function of the form $\phi_{X-\mu}(t) = \Psi(t^T \Sigma t)$, where Ψ is a $[0, \infty) \to \mathcal{R}$ function (characteristic generator) and $\Sigma \in \mathcal{R}^{n \times n}$ a symmetric positive definite matrix. If the density functions exist, it has the form

$$f(x) = |\Sigma|^{-\frac{1}{2}} g[(x-\mu)^T \Sigma^{-1} (x-\mu)], \quad x \in \mathcal{R}^n,$$

for some $[0, \infty) \to [0, \infty)$ function g (density generator).

Taking $g(y) = \frac{1}{2\pi} \exp\{-\frac{y}{2}\}$ gives the Gaussian distribution (Section 5.2.1) and $g(y) = \left(1 + \frac{ty}{\nu}\right)^{-\frac{2+\nu}{2}}$ leads to a Student's t distribution with ν degrees of freedom (Section 5.2.2).

Schmidt [7, Theorem 2.4α] shows that elliptical distributions are upper and lower tail dependent if the tail of their density generator is a regularly varying function with index $\alpha < -n/2$. A function g is called **regularly varying with index** α if for every $t > 0$

$$\lim_{x \to \infty} \frac{g(tx)}{g(x)} = t^{\alpha}.$$

Whether or not the generator being regularly varying is a necessary condition for tail dependence is still an open problem, but Schmidt [7, Theorem 2.4γ] proves that to have tail dependency the density generator g must be **O-regularly varying**, that is it must satisfy

$$0 < \liminf_{x \to \infty} \frac{g(tx)}{g(x)} \leq \limsup_{x \to \infty} \frac{g(tx)}{g(x)} < \infty,$$

for every $t \geq 1$.

5.2.1 Bivariate Gaussian copula

The bivariate Gaussian copula is defined as

$$C^{Ga}(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)),$$

where

$$\Phi_\rho(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{\frac{2\rho st - s^2 - t^2}{2(1-\rho^2)}} \, ds \, dt$$

and Φ denotes the standard normal CDF.

The Gaussian copula generates the joint standard normal distribution iff. $u = \Phi(x)$ and $v = \Phi(y)$, that is iff. the margins are standard normal.

Gaussian copulas have no tail dependency unless $\rho = 1$. This follows from Schmidt's [7] characterisation of tail dependent elliptical distributions, since the density generator for the bivariate Gaussian distribution ($\rho \neq 1$) is not O-regularly varying:

$$\lim_{x \to \infty} \frac{g(tx)}{g(x)} = \lim_{x \to \infty} \exp\{-\frac{1}{2} x(t-1)\} = 0, \quad t \geq 1.$$

5.2.2 Bivariate Student's t copula

Let t_ν denote the central univariate Student's t distribution function, with ν degrees of freedom:

$$t_\nu(x) = \int_{-\infty}^{x} \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\,\Gamma(\nu/2)} \left(1+\frac{s^2}{\nu}\right)^{-\frac{\nu+1}{2}} ds,$$

where Γ is Euler function and $t_{\rho,\nu}$, $\rho \in [0,1]$, the bivariate distribution corresponding to t_ν:

$$t_{\rho,\nu}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1+\frac{s^2+t^2-2\rho st}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} ds\,dt.$$

The bivariate Student's copula $T_{\rho,\nu}$ is defined as

$$T_{\rho,\nu}(u,z) = t_{\rho,\nu}(t_\nu^{-1}(u), t_\nu^{-1}(v)).$$

The generator for the Student's t is regularly varying:

$$\lim_{x\to\infty} \frac{g(tx)}{g(x)} = \lim_{x\to\infty} \left(1+\frac{tx}{\nu}\right)^{-\frac{2+\nu}{2}} \left(1+\frac{x}{\nu}\right)^{\frac{2+\nu}{2}} = \lim_{x\to\infty} \left(\frac{\nu+x}{\nu+tx}\right)^{\frac{2+\nu}{2}} = t^{-\frac{2+\nu}{2}}.$$

It follows that the Student's t distribution has tail dependence for all $\nu > 0$.

5.3 Archimedean copulas

Every continuous, decreasing, convex function $\phi : [0,1] \to [0,\infty)$ such that $\phi(1) = 0$ is a **generator** for an Archimedean copula. If furthermore $\phi(0) = +\infty$, then the generator is called **strict**. Parametric generators give rise to families of Archimedean copulas.

Define the pseudo-inverse of ϕ as

$$\phi^{[-1]} = \begin{cases} \phi^{-1}(u), & 0 \le u \le \phi(0), \\ 0, & \phi(0) \le u \le \infty. \end{cases}$$

In case of a strict generator, $\phi^{[-1]} = \phi^{-1}$ holds.

The function

$$C^A(u,v) = \phi^{[-1]}(\phi(u)+\phi(v)) \tag{5.1}$$

is a copula [2, Theorem 4.1.4] and is called the **Archimedean copula with generator ϕ**. The density of C^A is given by

$$c^A(u,v) = \frac{-\phi''(C(u,v))\phi'(u)\phi'(v)}{[\phi'(C(u,v))]^3}.$$

Table 5.1: *One-parameter Archimedean copulas. The families marked with * include C^-, $C^\perp$ and C^+.*

Name	$C_\theta(u,v)$	$\phi_\theta(t)$	$\theta \in$	τ	λ_L	λ_U
Clayton*	$\left(\max\{0, u^{-\theta}+v^{-\theta}-1\}\right)^{-\frac{1}{\theta}}$	$\frac{1}{\theta}(t^{-\theta}-1)$	$[-1,\infty)\backslash\{0\}$	$\frac{\theta}{\theta+2}$	$2^{-\frac{1}{\theta}}$	0
Gumbel-Hougaard	$\exp\left(-\left[(-\log u)^\theta+(-\log v)^\theta\right]^{\frac{1}{\theta}}\right)$	$(-\log t)^\theta$	$[1,\infty)$	$\frac{\theta-1}{\theta}$	0	$2-2^{\frac{1}{\theta}}$
Gumbel-Barnett	$uv\exp(-\theta\log u\log v)$	$\log(1-\theta\log t)$	$(0,1]$		0	0
Frank*	$-\frac{1}{\theta}\log\left(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$	$-\log\frac{e^{-\theta t}-1}{e^{-\theta}-1}$	$(-\infty,\infty)\backslash\{0\}$		0	0

5.3.1 One-parameter families

The Gumbel copula from example 4.12 is Archimedean with generator $\phi(u) = (-\log(u))^\theta$, $\theta \in [1,\infty)$. Some other examples are listed in table 5.1.

The Fréchet-Hoeffding lower bound C^- is Archimedean ($\phi(u) = 1-u$), whereas the Fréchet-Hoeffding upper bound is not. To see this, note that $\phi^{[-1]}$ is strictly decreasing on $[0,\phi(0)]$. Clearly, $2\phi(u) > \phi(u)$, so we have for the diagonal section of an Archimedean copula that

$$C^A(u,u) = \phi^{[-1]}(2\phi(u)) < \phi^{[-1]}(\phi(u)) = u. \tag{5.2}$$

As $C^+(u,u) = u$, inequality (5.2) implies that C^+ is not Archimedean.

Marshall and Olkin [8] show that if $\Lambda(\theta)$ is a distribution function with $\Lambda(0) = 0$ and Laplace transform

$$\psi(t) = \int_0^\infty e^{-\theta t} d\Lambda(\theta),$$

then $\phi = \psi^{-1}$ generates a strict Archimedean copula.

5.3.2 Two-parameter families

Nelsen [2] shows that if ϕ is a strict generator, then also $\phi(t^\alpha)$ (interior power family) and $[\phi(t)]^\beta$ (exterior power family) are strict generators for $\alpha \in (0,1]$ and $\beta \geq 1$. If ϕ is twice differentiable, then the interior power family is a strict generator for all $\alpha > 0$. Two-parameter families of Archimedean copulas can now be constructed by taking

$$\phi_{\alpha,\beta} = [\phi(t^\alpha)]^\beta$$

as the generator function.

For example, choosing $\phi(t) = \frac{1}{t} - 1$ gives $\phi_{\alpha,\beta} = (t^{-\alpha}-1)^\beta$ for $\alpha > 0$ and $\beta \geq 1$. This generates the family

$$C_{\alpha,\beta}(u,v) = \left\{\left[(u^{-\alpha}-1)^\beta + (v^{-\alpha}-1)^\beta\right]^{\frac{1}{\beta}} - 1\right\}.$$

For $\beta = 1$ this is (part of) the one-parameter Clayton family — see table 5.1.

5.3.3 Multivariate Archimedean copulas

This section extends the notion of an Archimedean copula to dimensions $n \geq 2$.

Kimberling [9] proves that if ϕ is a strict generator satisfying

$$(-1)^k \frac{d^k \phi^{-1}(t)}{dt^k} \geq 0 \quad \text{for all } t \in [0, \infty), \;\; k = 1, \ldots, n \tag{5.3}$$

then

$$C^A(u_1, \ldots, u_n) = \phi^{-1}(\phi(u_1) + \ldots + \phi(u_n))$$

is an n-copula.

For example, the generator $\phi_\theta(t) = t^{-\theta} - 1$ ($\theta > 0$) of the bivariate Clayton family has inverse $\phi_\theta^{-1}(t) = (1+t)^{-\frac{1}{\theta}}$ which is readily seen to satisfy (5.3). Thus,

$$C_\theta(u_1, \ldots, u_n) = \left(u_1^{-\theta} + \ldots + u_n^{-\theta} - n + 1\right)^{-\frac{1}{\theta}}$$

is a family of n-copulas.

It can be proven (see [10]) that Laplace transforms of distribution functions $\Lambda(\theta)$ satisfy (5.3) and $\Lambda(0) = 1$. The inverses of these transforms thus are a source of Archimedean n-copulas.

Archimedean n-copulas have practical restraints. To begin with, all k-margins are identical. Also, since there are usually only two parameters, Archimedean n-copulas are not very flexible to fit the n dimensional dependence structure. Furthermore, Archimedean copulas that have generators with complete monotonic inverse, are always more concordant than the product copula, i.e. they always model positive dependence.

There exist extensions of Archimedean copulas that have a number of mutually distinct bivariate margins. This is discussed in the next section.

5.4 Extension of Archimedean copulas

Generators of Archimedean copulas can be used to construct other (non-Archimedean) copulas. One such extension is discussed in Joe [11]. Copulas that are constructed in this way have the property of partial exchangeability, i.e. a number of bivariate margins are mutually distinct. We will only address the three dimensional case, but generalizations to higher dimensions are similar.

First, let ϕ be a strict generator and note that Archimedean copulas are associative:

$$\begin{aligned} C(u,v,w) &= \phi^{-1}(\phi(u) + \phi(v) + \phi(w)) \\ &= \phi^{-1}(\phi \circ \underbrace{\phi^{-1}(\phi(u) + \phi(v))}_{C(u,v)} + \phi(w)) \\ &= C(C(u,v), w). \end{aligned}$$

If we would choose the generator of the 'inner' and the 'outer' copula to be different, would the composition then still be a copula? In other words, for which functions ϕ, ψ is

$$C_\phi(C_\psi(u,v), w) \tag{5.4}$$

a copula? If it is, then the (1,2) bivariate margin is different from the (1,3) margin, but the (2,3) margin is equal to the (1,3) margin.

For $n = 1, 2, \ldots, \infty$, consider the following function classes:

$$\mathcal{L}_n = \left\{ \phi : [0, \phi) \to [0, 1] \,\middle|\, \phi(0) = 1, \phi(\infty) = 0, (-1)^k \frac{d^k \phi(t)}{dt^k} \geq 0 \text{ for all } t \in [0, \infty),\ k = 1, \ldots, n \right\},$$

$$\mathcal{L}_n^* = \left\{ \omega : [0, \infty) \to [0, \infty) \,\middle|\, \omega(0) = 0, \omega(\infty) = \infty, (-1)^{k-1} \frac{d^k \omega(t)}{dt^k} \geq 0 \text{ for all } t \in [0, \infty),\ k = 1, \ldots, n \right\},$$

Note that if $\phi^{-1} \in \mathcal{L}_1$, then ϕ is a strict generator for an Archimedean copula.

It turns out that if $\phi, \psi \in \mathcal{L}_1$ and $\phi \circ \psi^{-1} \in \mathcal{L}_\infty^*$, then (5.4) is a copula. For general n-copulas similar conditions exist. In the n-dimensional case, $n - 1$ of the $\frac{1}{2}n(n-1)$ bivariate margins are distinct.

6 The multivariate-multitemporal pricing model

The multivariate-multitemporal (multi-multi) model [12] is a pricing framework for multi-asset financial contracts where the payoff depends on a discrete number of observations. The basic building block is a multivariate Black-Scholes framework, i.e. a Gaussian copula and lognormal marginal distributions, resulting from modelling the forward price processes of the underlyings by correlated Brownian motions. To be consistent with single-asset derivatives prices, this model is then fitted to market prices by replacing the lognormal marginal distributions by market observed marginal densities. These are parametrized (terminal) densities resulting from a Displaced Diffusion (Section 6.1.1) or SABR (Section 6.1.2) process.

First we will discuss the case where the Gaussian terminal copula is taken from the multivariate Black-Scholes setup and only the marginal distributions are replaced by risk-neutral distributions (Section 6.2.1).

Next, in Section 6.2.2, not only the marginal distributions, but also the Gaussian copula will be replaced to allow for tail dependence. Since multi-asset derivatives in general are not liquid in the market, this copula will not be risk-neutral (like the implied marginal densities) but is calibrated to historical market data.

6.1 Marginal distributions

Unlike the Black-Scholes model whose marginal terminal densities are lognormal, the multi-multi model uses market implied marginal distributions (Appendix B.1). These distributions are inferred from vanilla option prices.

To determine the market implied densities, quotes are needed for call options at every possible strike. Such a continuum of prices is not available in the market. Instead of interpolating the available option prices to implicitly fix the marginal distribution, it is common to parametrize the implied volatility surface. In the standard Black-Scholes setup, specifying all implied volatilities completely determines the Black-implied distribution:

$$\mathbb{P}[F^T(T) < K] \stackrel{\text{(B.3)}}{=} \Phi(-d_2) + F^T(0)\,\sqrt{T}\,\phi(d_1)\,\frac{\partial \widehat{\sigma}}{\partial K}\,.$$

Here, $F^T(0)$ is the forward value of the underlying at maturity T seen from inception, $\widehat{\sigma}$ is the Black-implied volatility and d_1, d_2 are given by (A.11) and (A.12) respectively.

Sections 6.1.1 and 6.1.2 describe two volatility parametrizations. The corresponding terminal distributions have tails that are heavier than in the lognormal case, i.e. they incorporate volatility smile.

6.1.1 Displaced Diffusion

The displaced diffusion model assumes forward prices follow the dynamics

$$dF^T(t) = \mu F^T(t)\,dt + \sigma(t)\,[\beta S(t) + (1-\beta)F^T(0)]dW(t)\,. \tag{6.1}$$

Figure 6.1: *SABR implied volatility smile* ($\beta = 0.5, \sigma_0 = 1, \alpha = 0.4, \rho = -0.3$).

The parameter β takes real values and determines the slope of the skew. For $\beta = 0$ and $\beta = 1$ displaced diffusion corresponds to a normal and a lognormal process respectively. At time T the solution of the stochastic differential equation (6.1) is

$$F^T(T) = \frac{F^T(0)}{\beta}\left[\exp\left(-\frac{1}{2}\beta^2\int_0^T \sigma^2(s)\,ds \,+\, \beta\int_0^T \sigma(s)\,dW(s)\right) - (1-\beta)\right]. \tag{6.2}$$

The distribution of $\int_0^T \sigma(s)dW(s)$ is known to be $\mathcal{N}(0,\widehat{\sigma})$ where $\widehat{\sigma}^2 = \frac{1}{T}\int_0^T \sigma(s)^2\,ds$. It follows that

$$\mathbb{P}[S(T) < K] \;=\; \Phi\left(-\frac{\log\frac{F^T(0)}{\beta K-(1-\beta)F^T(0)} - \frac{1}{2}\beta^2\int_0^T \sigma^2(s)\,ds}{\beta\left(\int_0^T \sigma^2(s)\,ds\right)^{\frac{1}{2}}}\right).$$

6.1.2 SABR

The stochastic-$\alpha\beta\rho$ (SABR) model, originally proposed by Hagan et al. [13], models forward price and volatility as two correlated processes:

$$\begin{cases} dF^T(t) &=& \sigma(t)\,F^T(t)^\beta\,dW_1(t)\,, \\ d\sigma(t) &=& \alpha\,\sigma(t)\,dW_2(t)\,, \qquad\qquad \sigma(0) = \sigma_0\,, \\ dW_1\,dW_2 &=& \rho\,dt\,. \end{cases} \tag{6.3}$$

The correlation ρ between the forward and the volatility process takes values on $[-1,1]$, the instantaneous volatility σ and the 'volatility of volatility' α are positive and β is any real number, though generally lies in $[0,1]$. These dynamics produce a 'smile' in the implied volatility curve of the terminal distribution (Figure 6.1). Prices of European options can be approximated using perturbation techniques [13, Appendix A]. From this an expression for the implied volatility can be derived:

$$\widehat{\sigma}(F^T(0),K,T,\sigma,\alpha,\beta,\rho) = \sigma_0\,\frac{1+\eta T}{\sqrt{g}\,h}\,\frac{z}{\chi}\,, \tag{6.4}$$

where

$$
\begin{aligned}
g &:= (F^T(0)\,K)^{1-\beta}\,, \\
x &:= \log(F^T(0)/K)\,, \\
h &:= 1 + \frac{(1-\beta)^2}{24}x^2 + \frac{(1-\beta)^4}{1920}x^4\,, \\
\eta &:= \frac{(1-\beta)^2}{24}\frac{\sigma_0^2}{g} + \frac{\sigma_0\alpha\beta\rho}{4\sqrt{g}} + \frac{2-3\rho^2}{24}\alpha^2\,, \\
z &:= \frac{\alpha}{\sigma_0}x\sqrt{g}\,, \\
\chi &:= \log\frac{z-\rho+\sqrt{1-2\rho z+z^2}}{1-\rho}\,.
\end{aligned}
$$

6.2 Dependence

If we are pricing a hybrid contract over n underlyings $S_1, \ldots, S_n$ where the payoff depends on observations at m timepoints $t_1, \ldots, t_m$, we need a copula specifying the dependence structure between the terminal distributions $S_1(t_1), \ldots, S_n(t_1), \ldots, S_1(t_m), \ldots, S_n(t_m)$. We will refer to this as **terminal dependence**.

For calibration one usually looks at **instantaneous dependence** — i.e. dependence between periodic returns — instead of dependence between terminal levels. The reason for this is that consecutive asset prices are not independent, whereas autocorrelation in daily returns usually is low — see for example Table 6.1. Autocorrelation between levels (not listed in the table) generally is near unity. The presence of autocorrelation in data series obstructs the calibration of the copula since the maximum likelihood approach assumes the observations to be independent (Section 8.1).

Table 6.1: *Autocorrelation in daily and monthly log-returns.*

	Autocorrelation in log-returns	
	Daily	**Monthly**
SP500	0.005	-0.098
NIKKEI 225	−0.016	0.062
NASDAQ	0.056	0.124
Corn	0.004	0.100
Wheat	−0.044	−0.077

If we calibrate a copula to daily or monthly returns, we need to 'transform' the instantaneous dependence described by this copula into terminal dependence. In case of a multivariate Black-Scholes model with correlated Brownian motions driving the forward price processes, there is an analytic way to calculate the terminal covariances (Section 6.2.1). If we are to use a general copula to allow for tail dependence, we have to calculate the convolution of daily or monthly increments by means of Monte Carlo simulation (Section 6.2.2).

Figure 6.2: *Correlation matrix used in the multi-multi model. The cross-time correlations (▲) are given by equation (6.7), the cross-asset correlations (○) by (6.8).*

	$S_1(t_1)$	$\dots$	$S_1(t_m)$	$S_2(t_1)$	$\dots$	$S_2(t_m)$	$\dots$	$S_n(t_1)$	$\dots$	$S_n(t_m)$
$S_1(t_1)$	1	▲	▲	○	○	○		○	○	○
$\vdots$		$\ddots$	▲	○	○	○		○	○	○
$S_1(t_m)$			1	○	○	○		○	○	○
$S_2(t_1)$				1	▲	▲		○	○	○
$\vdots$					$\ddots$	▲		○	○	○
$S_2(t_m)$						1		○	○	○
$\vdots$							$\ddots$			
$S_n(t_1)$								1	▲	▲
$\vdots$									$\ddots$	▲
$S_n(t_m)$										1

6.2.1 Normal copula

The multi-multi framework basically is a multivariate Black-Scholes model where the marginal distributions are replaced to incorporate volatility smile and to be consistent with single-asset derivatives prices. The covariance matrix of the Gaussian copula however is taken from the multivariate Black-Scholes case, i.e. as if the marginal distributions were lognormal. The Brownian motions driving the spot price processes thus are assumed to follow the dynamics

$$\begin{cases} dx(t) & = & \sigma_x(t)\, dW_x(t)\,, \\ dy(t) & = & \sigma_y(t)\, dW_y(t)\,, \\ dW_x(t)\, dW_y(t) & = & \rho\, dt\,, \end{cases} \tag{6.5}$$

where σ_x, σ_y are the instantaneous volatilities of the Brownian motions. The instantaneous correlation ρ can be calculated via

$$\rho = 2\sin(\,\frac{\pi}{6}\,\rho_S\,)$$

where ρ_S is Spearman's rank correlation coefficient for the logarithm of the returns of the spot prices.

$$\begin{aligned} \mathrm{Var}[x(t)] &= \mathbb{E}\left[x(t)^2\right] - \underbrace{\left(\mathbb{E}\left[x(t)\right]\right)^2}_{=0} \\ &= \mathbb{E}\left[\left(\int_0^t \sigma_x(s)\, dW_x(s)\right)^2\right] \\ &\overset{\text{Itō isometry}}{=} \int_0^t |\,\sigma_x(s)\,|^2\, ds \;:=\; t\,\overline{\sigma}_x^2(t) \end{aligned} \tag{6.6}$$

The variance per unit time $\overline{\sigma}_x^2$ is chosen to match the at-the-money implied volatilities at $t_1, \dots, t_m$.

For $t \leq s$ the covariance of $x(t)$ and $x(s)$ is

$$
\begin{aligned}
\mathrm{Cov}[x(t), x(s)] &= \mathbb{E}\left[\, x(t)\, x(s)\,\right] - \underbrace{\mathbb{E}\left[x(t)\right]\mathbb{E}\left[x(s)\right]}_{=0} \\
&= \mathbb{E}\left[\int_0^t \sigma_x(u)\, dW_x(u) \int_0^s \sigma_x(u)\, dW_x(u)\right] \\
&= \mathbb{E}\left[\left(\int_0^t \sigma_x(u)\, dW_x(u)\right)^2 + \int_0^t \sigma_x(u)\, dW_x(u) \int_t^s \sigma_x(u)\, dW(u)\right] \\
&\overset{\substack{\text{Indep.}\\ \text{increments}}}{=} \mathbb{E}\left[x(t)^2\right] \overset{(6.6)}{=} t\, \overline{\sigma}_x^2(t)\,.
\end{aligned}
$$

Dividing by the product of the standard deviations gives the cross-time correlation

$$
\mathrm{Corr}(x(t), x(s)) = \frac{\mathrm{Cov}[x(t), x(s)]}{\sqrt{\mathrm{Var}[x(t)]\,\mathrm{Var}[x(s)]}} = \frac{\overline{\sigma}_x^2(t)\, t}{\overline{\sigma}_x(t)\, \sqrt{t}\, \overline{\sigma}_x(s)\, \sqrt{s}} = \frac{\overline{\sigma}_x(t)}{\overline{\sigma}_x(s)}\, \frac{\sqrt{t}}{\sqrt{s}}\,. \tag{6.7}
$$

Equation (6.6) must hold for the at-the-money volatilities at $t = t_1, \ldots, t_m$, but this constraint is satisfied by infinitely many instantaneous volatility functions σ_x. By assuming some term structure for the instantaneous volatility (e.g. piecewise constant) we can fix σ_x, σ_y and calculate the cross-asset covariance

$$
\begin{aligned}
\mathrm{Cov}[x(t), y(s)] &= \mathbb{E}\left[\, x(t) y(s)\,\right] - \underbrace{\mathbb{E}\left[x(t)\right]\mathbb{E}\left[y(s)\right]}_{=0} \\
&= \mathbb{E}\left[\int_0^t \sigma_x(u)\, dW_x(u) \int_0^t \sigma_y(u)\, dW_y(u)\right] \\
&\quad + \mathbb{E}\left[\int_0^t \sigma_x(u)\, dW_x(u) \int_t^s \sigma_y(u)\, dW_y(u)\right] \\
&\overset{\substack{\text{Indep.}\\ \text{increments}}}{=} \mathbb{E}\left[\int_0^t \sigma_x(u)\, dW_x(u) \int_0^t \sigma_y(u)\, dW_y(u)\right] \\
&= \int_0^t \rho\, \sigma_x(u)\, \sigma_y(u)\, du \;:=\; t\, \overline{\sigma}_{xy}^2(t)\,.
\end{aligned}
$$

Therefore,

$$
\mathrm{Corr}(x(t), y(s)) = \frac{\overline{\sigma}_{xy}^2(t)}{\overline{\sigma}_x(t)\, \overline{\sigma}_y(s)}\, \frac{\sqrt{t}}{\sqrt{s}}\,. \tag{6.8}
$$

Equations (6.7) and (6.8) together completely specify the covariance matrix of the normal copula.

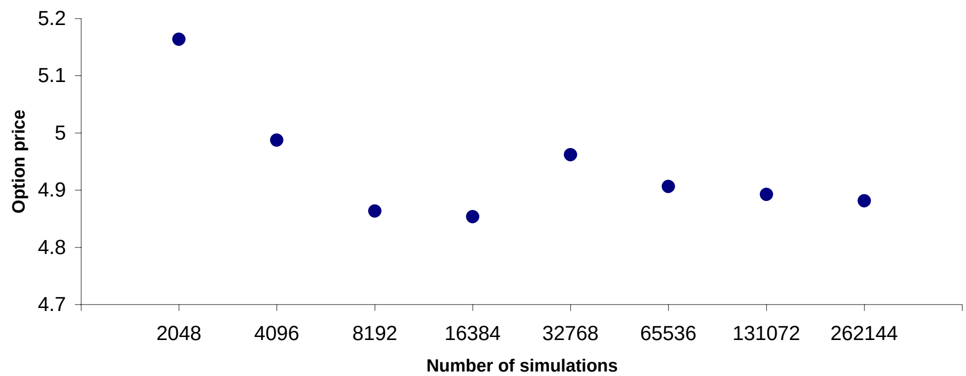


Figure 6.3: *Convergence of a worst-of option price for the multi-multi model with a Clayton survival copula calibrated to daily returns.*

6.2.2 Alternative copula

In the preceding section the copula modelling the dependence between underlyings was assumed to be Gaussian. We will now lift this assumption and extend the multi-multi model for use with a general copula, but limit ourselves to options where the payoff depends on observations at maturity only. In case of the Gaussian copula the covariance matrix was taken from the multivariate Black-Scholes model. For a general copula, which is not parametrized in terms of covariances, this approach cannot be used.

The maximum likelihood method (Section 8.1) that is used to calibrate the copula assumes consecutive observations in the historical data set to be independent. Empirical evidence indicates that price levels are not independent. Therefore we will focus on (log-)returns instead.

The **copula**, being **calibrated to returns**, captures instantaneous dependence between underlyings, that is, dependence between price changes on a daily, weekly or monthly timescale. By compounding several of those price changes we can construct returns for larger timescales, for instance the return to maturity. Up to a multiplication by the (non-stochastic) initial price level this is the same as the terminal level whose dependence structure we are interested in.

Since we are now modelling returns rather than levels, we should also specify the **marginal distribution of the returns**. To be consistent with liquid single-asset contracts, we will derive the marginal return densities from the same processes we previously used to generate terminal univariate price distributions:

Lognormal case In case of lognormal margins and under risk neutrality the forwards $F_1, \ldots, F_n$ with maturity t_m of the underlyings follow the process

$$dF_j(t) \;=\; \sigma_j(t)\, F_j(t)\, dW_j(t), \qquad j = 1, \ldots, n\,. \tag{6.9}$$

Suppose we have a model for the dependence between Δt-periodic returns. Partition the interval $[0, t_m]$, into subintervals of length Δt. We can simulate observations from the forwards at maturity t_m using the solution of (6.9): for $k = 0, \Delta t, 2\Delta t, \ldots, t_m/\Delta t - 1$ calculate

$$\log \frac{F_j(\,(k+1)\Delta t\,)}{F_j(\,k\Delta t\,)} = -\frac{1}{2}\Delta t\,(\sigma_j)^2 \;+\; \sqrt{\Delta t}\,\sigma_j\,\Phi^{-1}(u_j^k)\,. \tag{6.10}$$

The u_j^k are independent and uniformly distributed for fixed j. For fixed k, their joint distribution is determined by the choice of the copula.

In case of a normal copula, this procedure is equivalent to the approach described in Section 6.2.1 with lognormal margins.

Displaced diffusion case If the underlying follows a displaced diffusion process, formula (6.10) becomes

$$\log \frac{F_j(\,(k+1)\Delta t\,)}{F_j(\,k\Delta t\,)} = -\log\beta \;-\; \frac{1}{2}\beta^2\Delta t\,(\sigma_j)^2 \;+\; \beta\,\sqrt{\Delta t}\,\sigma_j\,\Phi^{-1}(u_j^k)\,. \tag{6.11}$$

By adding up the joint distributions of price increments over small time periods we arrive at the distribution of price changes on the time interval to maturity. This essentially means calculating a convolution. We will approximate this convolution by Monte Carlo simulation. Figure 6.3 shows that it roughly takes 100000 simulatons for the option price to converge.

Summarizing, to calculate the option price we proceed as follows:

Algorithm 6.1

1. *Calibrate a copula to historical* Δt *periodical returns (Section 8).*
2. *For each* k *in* $[0, t_m)$ *step size* Δt*:*
   (i) Simulate (Section 7) an observation $(u_1^k, \ldots, u_n^k)$ *from the copula obtained in step 1.*
   (ii) Update the forwards $F_1((k+1)\Delta t), \ldots, F_n((k+1)\Delta t)$ *using (6.10) or (6.11).*
3. *Calculate the option price which is a function of the spot prices* $F_1(t_m), \ldots, F_n(t_m)$ *of the underlyings at maturity.*
4. *Repeat steps 2 and 3 and average the option price.*

7 Simulation

This section covers methods for sampling from Archimedean copulas. Marshall and Olkin's approach (Section 7.2) generates samples from general n-dimensional Archimedean copulas but it requires a draw from the distribution corresponding to the inverse Laplace transformation of the generator of the copula. Alternatively, the conditional sampling approach (Section 7.1) does not require such a draw, but is not applicable for all Archimedean copulas (for instance not for the Gumbel copula) and sometimes involves equations that have to be solved numerically (like the n-Frank copula which requires solving an $n-1$ dimensional polynomial equation).

7.1 Conditional sampling

In the conditional sampling approach, the components of the realization of the copula are generated one by one. The first component is drawn from a uniform distribution, the second is based on the copula conditional on the first draw and so on. Examples 7.1 and 7.2 demonstrate this procedure for Clayton and Frank copulas.

Adopting the notation

$$C_k(u_1,\dots,u_k) := C(u_1,\dots,u_k,1,\dots,1)\,, \qquad 1 \le k \le n\,,$$

the conditional distribution of the k-th component given the preceding ones is

$$\begin{aligned}
\mathbb{P}[U_k \le u_k \,|\, U_1 = u_1,\dots,U_{k-1}=u_{k-1}] &= \frac{\mathbb{P}[U_k \le u_k, U_1 = u_1,\dots,U_{k-1}=u_{k-1}]}{\mathbb{P}[U_1 = u_1,\dots,U_{k-1}=u_{k-1}]} \\
&= \frac{\partial^{k-1} C_k(u_1,\dots,u_k)/(\partial u_1 \dots \partial u_{k-1})}{\partial^{k-1} C_{k-1}(u_1,\dots,u_{k-1})/(\partial u_1 \dots \partial u_{k-1})}\,.
\end{aligned}$$

Nelsen [2] therefore proposes the following algorithm:

- Simulate a vector $(v_1,\dots,v_n)$ of uniform random variates.
- Put $\begin{cases} v_1 = u_1 \\ v_2 = \mathbb{P}[U_2 \le u_2 \,|\, U_1 = u_1] \\ \qquad\vdots \\ v_k = \mathbb{P}[U_n \le u_n \,|\, U_1 = u_1,\dots,U_{n-1}=u_{n-1}] \end{cases}$
- The solution $(u_1,\dots,u_n)$ of this system is a realization of the copula.

Cherubini et al. [5] show that for Archimedean copulas

$$\mathbb{P}[U_k \le u_k \,|\, U_1 = u_1,\dots,U_{k-1}=u_{k-1}] = \frac{\frac{\partial^{k-1}\phi^{-1}}{\partial t^{k-1}}(\,\phi(u_1)+\dots+\phi(u_k)\,)}{\frac{\partial^{k-1}\phi^{-1}}{\partial t^{k-1}}(\,\phi(u_1)+\dots+\phi(u_{k-1})\,)}\,. \tag{7.1}$$

The following algorithms for generating Clayton and Frank variates are stated for the bivariate case, but can easily be generalized to the multidimensional case — see [5, p. 185].

Example 7.1 (Clayton copula simulation) *The generator of the Clayton copula is* $\phi(u) = u^{-\theta}-1$ *with inverse* $\phi^{-1}(t) = (t+1)^{-\frac{1}{\theta}}$ *and* $\partial_t \phi^{-1} = \frac{1}{\theta}(t+1)^{-\frac{1+\theta}{\theta}}$. *Draw* v_1, v_2 *from a uniform distribution on [0,1] and put*

$$\begin{cases} v_1 = u_1 \ , \\ \\ v_2 \overset{(7.1)}{=} \frac{\partial_t \phi^{-1}(\phi(u_1)+\phi(u_2))}{\partial_t \phi^{-1}(\phi(u_1))} = \left(\frac{u_1^{-\theta}+u_2^{-\theta}-1}{u_1^{-\theta}}\right)^{-\frac{1+\theta}{\theta}} \end{cases} .$$

Solving for u_2 *gives*

$$u_2 = \left(u_1^{-\theta}\left(v_2^{-\frac{\theta}{1+\theta}} - 1\right) + 1\right)^{-\frac{1}{\theta}} .$$

Figure 7.1 shows 700 *simulations from a Clayton copula with parameter* 5.

Example 7.2 (Frank copula simulation) *The generator of the Frank copula is* $\phi(u) = \log \frac{e^{-\theta u}-1}{e^{-\theta}-1}$ *with inverse* $\phi^{-1}(t) = -\frac{1}{\theta}\log(1+e^t(e^{-\theta}-1))$ *and* $\partial_t \phi^{-1} = -\frac{1}{\theta}\frac{e^t(e^{-\theta}-1)}{1+e^t(e^{-\theta}-1)}$. *Draw* v_1, v_2 *from a uniform distribution on [0,1] and put*

$$\begin{cases} v_1 = u_1 \ , \\ \\ v_2 \overset{(7.1)}{=} \frac{\partial_t \phi^{-1}(\phi(u_1)+\phi(u_2))}{\partial_t \phi^{-1}(\phi(u_1))} = \frac{e^{-\theta u_2}-1}{e^{-\theta}-1+(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)} \end{cases} .$$

Solving for u_2 *gives*

$$u_2 = -\frac{1}{\theta}\log\left(1 + \frac{v_2(1-e^{-\theta})}{v_2(e^{-\theta u_1}-1) - e^{-\theta u_1}}\right) .$$

Figure 7.2 shows 700 *simulations from a Frank copula with parameter* -10.

7.2 Marshall and Olkin's method

This section closely follows Marshall and Olkin [8].

Let F, G be two *univariate* distributions, then

$$H(x) := \int F(x)^\gamma dG(\gamma) \tag{7.2}$$

is also a distribution function — it is the mixture of distributions F^γ with weights determined by G. Recall that

$$\tau(x) = \int e^{-x\gamma} dG(\gamma)$$

is the Laplace transform of G. It follows that

$$\tau(-\log F(x)) = H(x) \, ,$$

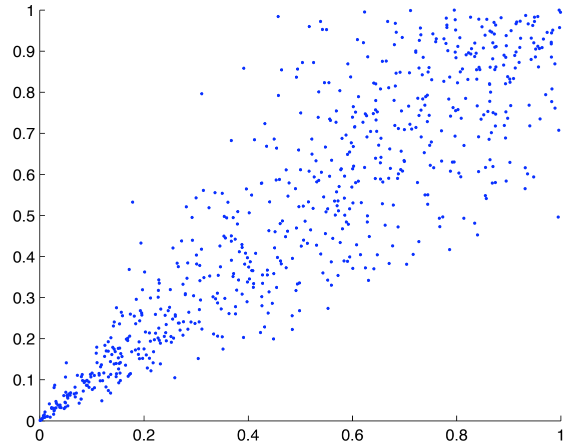


Figure 7.1: *Scatterplot of Clayton(5) copula.*

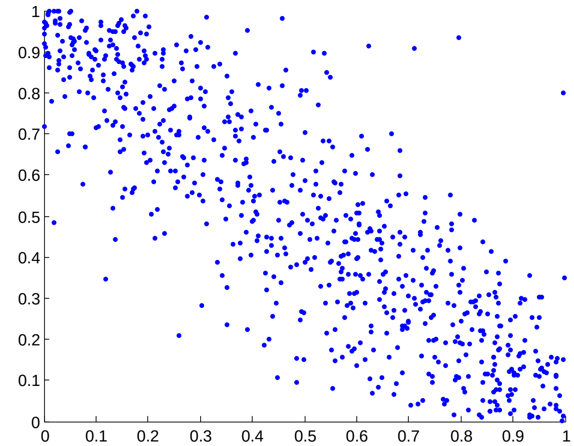


Figure 7.2: *Scatterplot of Frank(−10) copula.*

or equivalently,

$$F(x) = \exp(-\,\tau^{-1}H(x)\,)\,, \tag{7.3}$$

i.e. given H and G, there always exists a distribution function F such that (7.2) holds.

Now let G be a *bivariate* distribution having margins G_1, G_2. The margins H_1, H_2 of

$$H(x_1, x_2) = \iint F_1(x_1)^{\gamma_1} F_2(x_2)^{\gamma_2}\, dG(\gamma_1, \gamma_2) \tag{7.4}$$

are given by $H_i(x) = \int F_i^{\gamma_i}(x)\, dG_i(\gamma_i)$. Conversely, if

$$F_i(x) = \exp(-\tau_i^{-1} H_i(x))\,, \tag{7.5}$$

where τ_i is the Laplace transform of G_i, then H given by (7.4) is a bivariate distribution with margins H_1 and H_2.

Consider the special case where G is the Fréchet-Hoeffding upper bound and the margins G_i, $1 \leq i \leq n$, are equal:

$$G(\gamma_1, \ldots, \gamma_n) \;=\; C^+(G_1(\gamma_1), \ldots, G_1(\gamma_n)) \;=\; \min_{1 \leq i \leq n} G_1(\gamma_i)\,.$$

In this case the n-dimensional version of (7.4) reduces to

$$\begin{aligned} H(x_1, \ldots, x_n) &= \int F_1(x_1)^{\gamma} \ldots F_n(x_n)^{\gamma}\, dG_1(\gamma) \\ &= \tau_1(-\log[F_1(x_1) \ldots F_n(x_n)]\,) \\ &= \tau_1(-\log F_1(x_1) - \ldots - \log F_n(x_n)) \\ &\overset{(7.3)}{=} \tau_1\,[\,\tau_1^{-1}(\,H_1(x_1)\,) + \ldots + \tau_1^{-1}(\,H_n(x_n)\,)\,] \\ &= \phi^{-1}\,[\,\phi(\,H_1(x_1)\,) + \ldots + \phi(\,H_n(x_n)\,)\,]\,, \end{aligned} \tag{7.6}$$

where $\phi := \tau_1^{-1}$ is the inverse Laplace transform of G_1. This shows that Archimedean copulas can be thought of as mixture distributions. To simulate observations from (7.6) one can thus first determine the mixture component to be drawn from and then generate a vector from that distribution.

To determine the mixture component to be drawn from, simulate a number γ from G, the distribution whose Laplace transform is the inverse of ϕ. Next, generate a vector $(u_1, \ldots, u_n)$ of numbers from a uniform distribution over [0,1] and put

$$F_i(x)^\gamma = u_i, \qquad 1 \le i \le n\,.$$

From (7.5) it follows that the ranks are given by

$$H_i(x) = \tau(-\frac{1}{\gamma}\log u_i), \qquad 1 \le i \le n\,. \tag{7.7}$$

Example 7.3 (Gumbel copula simulation) *The generator $\phi(t) = (-\log t)^\theta$ of the Gumbel copula with inverse $\tau = \exp(-s^{\frac{1}{\theta}})$ is the Laplace transform of a stable distribution (Section B.2) with $\alpha = \frac{1}{\theta}$, $\beta = 1$ and both scale and location parameter 0. Chambers et al. [14] give the following algorithm to sample from this distribution:*

- *Draw a standard exponential variable W, i.e. $W = -\log(V)$ where V is uniform on $[0,1]$.*
- *Draw a variable Z from a uniform distribution on $[0,\pi]$.*
- *Set $\gamma = \frac{\sin((1-\frac{1}{\theta})Z)^{\theta-1}\sin(\frac{Z}{\theta})}{W^{\theta-1}\sin(Z)^\theta}$. This variable follows a Stable($\frac{1}{\theta}, 1$) distribution.*

Given a vector $(u_1, \ldots, u_n)$ of numbers from a uniform distribution over $[0,1]$ and using (7.7) one obtains the Gumbel realizations

$$H_i(x) = \exp(-(-\frac{1}{\gamma}\log u_i)^{\frac{1}{\theta}})\,, \qquad 1 \le i \le n\,.$$

Figure 7.3 shows 700 simulations from a Gumbel copula with parameter 2.5.

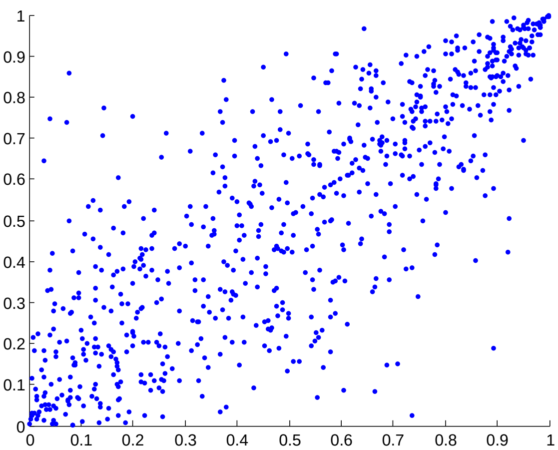


Figure 7.3: *Scatterplot of Gumbel*(2.5) *copula.*

8 Calibration of copulas to market data

This section is concerned with the question which member of a parametric family of copulas fits best to a given set of market data.

Consider a stochastic process $\{Y_t,\ t = 1, 2, \ldots\}$ taking values in $\mathcal{R}^n$. Our data consists in a realisation $\{(x_{1t}, \ldots, x_{nt}) : t = 1, \ldots, T\}$ of the vectors $\{Y_t,\ t = 1, \ldots, T\}$.

8.1 Maximum likelihood method

Let $X_i \sim F_i$, $1 \le i \le n$, be random variables with joint distribution H. From the multidimensional version of Sklar's theorem we know there exists a copula C such that

$$H(u_1, \ldots, u_n) = C(F(u_1), \ldots, F(u_n)).$$

Differentiating this expression with respect to $u_1, u_2, \ldots, u_n$ sequentially yields the **canonical representation**

$$h(u_1, \ldots, u_n) = c(F_1(u_1), \ldots, F_n(u_n)) \prod_{i=1}^{n} f_i(u_i), \tag{8.1}$$

where c is the copula density.

The maximum likelihood method implies choosing C and $F_1, \ldots, F_n$ such that the probability of observing the data set is maximal. The data set is assumed to consist of independent observations.

The possible choices for the copula and the margins are unlimited, or, in the words of Cherubini et al. [5], "*copulas allow a double infinity of degrees of freedom*". Therefore we usually restrict ourselves to certain classes of functions, parametrized by some vector $\theta \in \Theta \subset \mathcal{R}^n$.

We should thus find $\theta \in \Theta$ that maximizes the likelihood

$$l(\theta) := \prod_{t=1}^{T} \left(c(F_1(x_{1t}), \ldots, F_n(x_{nt});\ \theta) \prod_{i=1}^{n} f_i(x_{it};\ \theta) \right). \tag{8.2}$$

This θ also maximimes the log-likelihood

$$\log l(\theta) = \sum_{t=1}^{T} \log c(F_1(x_{1t}), \ldots, F_n(x_{nt});\ \theta) + \sum_{t=1}^{T} \sum_{i=1}^{n} \log f_i(x_{it};\ \theta). \tag{8.3}$$

The latter expression is often computationally more convenient. The vector θ that maximizes $l(\theta)$ is called the **maximum likelihood estimator** (MLE):

$$\theta_{\text{MLE}} := \underset{\theta \in \Theta}{\text{argmax}}\ l(\theta).$$

If $\partial l(\theta)/\partial\theta$ exists, then the solutions of

$$\frac{\partial l(\theta)}{\partial \theta} = 0$$

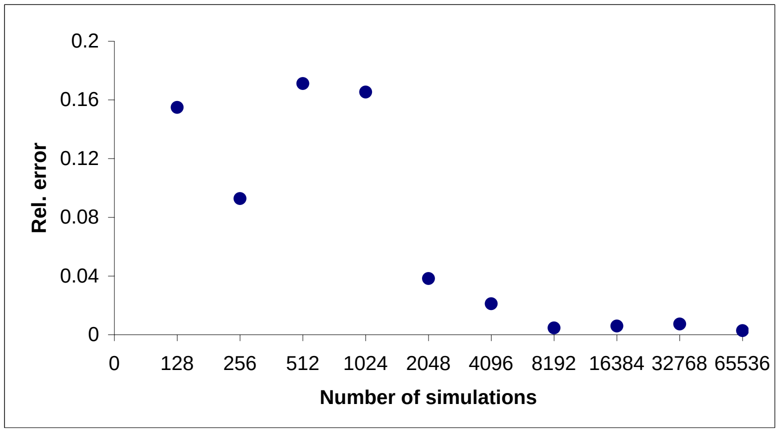


Figure 8.1: *Maximum likelihood estimation applied to a Clayton(1.4) copula for different sample sizes.*

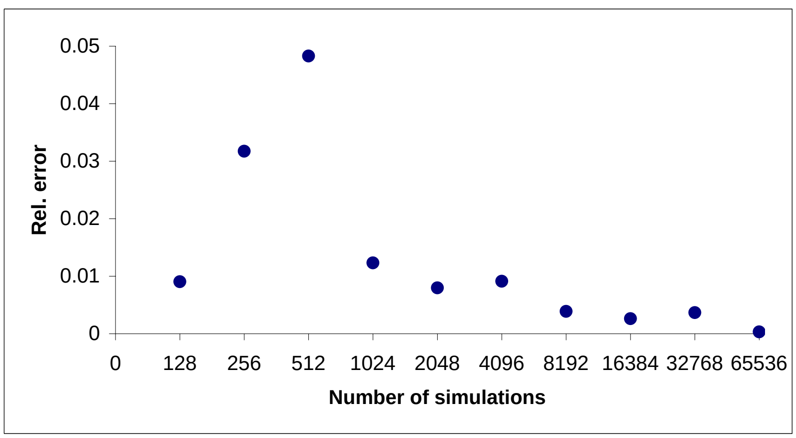


Figure 8.2: *Maximum likelihood estimation applied to a Gumbel(1.2) copula for different sample sizes.*

are possible candidates for θ_{MLE}. But these solutions can also be local maxima, minima or inflection points. On the other hand, maxima can occur at the boundary of Θ (or if $||\theta|| \to \infty$), in discontinuity points and in points where the likelihood is not differentiable.

For joint distributions satisfying some regularity conditions, it can be shown (Shao [15]) that if the sample size increases, the subsequent MLEs converge to a limit. This property is called consistency.

Figures 8.1 and 8.2 show relative errors in the estimation of the parameter of a Clayton and a Gumbel copula for different sample sizes. The samples were generated via the methods described in Section 7. These experiments show that maximum likelihood estimation applied to a sample containing less than 1000 observations can be expected to lead to relative errors that exceed 5%.

8.2 IFM method

The log-likelihood (8.3) consists of two positive parts. Joe and Xu [16] proposed the set of parameters θ to be estimated in two steps: first the margins' parameters and then the copulas'. By doing so, the computational cost of finding the optimal set of parameters reduces significantly. This approach is called the Inference for the Margins (IFM) method.

$$\begin{aligned} \theta_1 &= \underset{\theta_2}{\text{argmax}} \sum_{t=1}^{T} \sum_{i=1}^{n} \log f_i(x_{it};\ \theta_1) \\ \theta_2 &= \underset{\theta_1}{\text{argmax}} \sum_{t=1}^{T} \log c(F_1(x_{1t}), \ldots, F_n(x_{nt});\ \theta_1, \theta_2) \end{aligned}$$

Set $\theta_{\text{IFM}} := (\theta_1, \theta_2)$ to be the IFM estimator. In general, this estimator will be different from the MLE, but it can be shown that the IFM estimator is consistent.

8.3 CML method

In the Canonical Maximum Likelihood (CML) method first the margins are estimated using empirical distributions $\hat{F}_1, \ldots, \hat{F}_n$. Then, the copula parameters are estimated using an ML approach:

$$\theta_{\text{CML}} := \underset{\theta}{\text{argmax}} \sum_{t=1}^{T} \log c(\hat{F}_1(x_{1t}), \ldots, \hat{F}_n(x_{nt});\ \theta).$$

8.4 Expectation Maximization algorithm

Even though the number of parameters to be optimized simultaneously is reduced by the IFM or CML approach, finding the maximum likelihood estimator can still be hard for a mixture of copulas. The Expectation Maximization (EM) algorithm provides a way to find the weights and the parameters that maximize the likelihood of mixture distributions.

In short the EM algorithm proceeds as follows. First a lower bound to the log-likelihood is constructed. This lower bound can be interpreted as an expected value. Given the current parameter estimate it finds the distribution for which the expected value is maximal. Fixing the distribution, the algorithm then finds the parameters that maximize the lower bound. This will turn out to be surprisingly easy as the optimization over the weights and each of the component densities can be done separately. Repeating this procedure yields a sequence of parameter estimates with monotonically increasing likelihood.

We will now describe the EM algorithm in more detail, based on [17] and [18]. Also, convergence will be discussed and we will compare EM to alternative numerical methods.

Given a dataset $X = (x_1, \ldots, x_m)$ that is assumed to be generated from a mixture distribution with density

$$\mathbb{P}[x|B, \Theta] = \sum_{i=1}^{n} \beta_i \, p_i(\, x \,|\, \theta_i \,) \;, \qquad \sum_{i=1}^{n} \beta_i = 1 \;, \tag{8.4}$$

where $\Theta = (\theta_1, \dots, \theta_n)$, $B = (\beta_1, \dots, \beta_n)$ and $p_1, \dots, p_n$ are the component density functions, consider the problem of finding (B, Θ) maximizing the log-likelihood

$$\ell(B,\Theta|X) \;:=\; \log \mathbb{P}[X|B,\Theta] \;=\; \sum_{j=1}^{m} \log\left(\sum_{i=1}^{n} \beta_i\, p_i(\, x_j \,|\, \theta_i\,)\right). \tag{8.5}$$

We cannot observe the random vector $Y = (y_1, \dots, y_m) \in \{1, \dots, n\}^m$ whose values represent the components by which each data item was generated. The EM algorithm therefore finds parameters (B, Θ) that maximize the log-likelihood of the dataset X :

$$\ell(B,\Theta|X) \;=\; \log \mathbb{P}[X|B,\Theta] \;=\; \log \sum_{Y\in\{1,\dots,n\}^m} \mathbb{P}[X,Y|B,\Theta] \tag{8.6}$$

Denote by f any discrete probability distribution of Y over $\{1, \dots, n\}^m$. Jensen's inequality[1] provides a lower bound to the log-likelihood (8.6):

$$\underbrace{\sum_{Y\in\{1,\dots,n\}^m} f(Y)\,\log\frac{\mathbb{P}[X,Y|B,\Theta]}{f(Y)}}_{:=\,Q^f(B,\Theta)} \;\le\; \underbrace{\log \sum_{Y\in\{1,\dots,n\}^m} f(Y)\frac{\mathbb{P}[X,Y|B,\Theta]}{f(Y)}}_{=\,\ell(B,\Theta|X)}. \tag{8.7}$$

The next step is to determine the distribution f^t for which the lower bound $Q^{f^t}(B,\Theta)$ is maximal in the current parameter estimate (B^t, Θ^t). To ensure $\sum_{Y\in\{1,\dots,n\}^m} f^t(Y) \;=\; 1$ a Lagrange multiplier λ is used. The objective function becomes

$$G(f^t) := \lambda\left[1 - \sum_{Y\in\{1,\dots,n\}^m} f^t(Y)\right] + \sum_{Y\in\{1,\dots,n\}^m} f^t(Y)\,\log\mathbb{P}[X,Y|B^t,\Theta^t] - \sum_{Y\in\{1,\dots,n\}^m} f^t(Y)\,\log f^t(Y)\,.$$

As f^t is a discrete distribution which is completely specified by the n^m values $f^t(Y)$, $Y \in \{1, \dots, n\}^m$, we can set the derivative of G to every one of these values equal to zero:

$$0 \;=\; \frac{\partial G}{\partial f^t(Y)} \;=\; -\lambda \;+\; \log\mathbb{P}[X,Y|B^t,\Theta^t] \;-\; \log f^t(Y) - 1\,. \tag{8.8}$$

Equation (8.8) is equivalent to

$$f^t(Y) = e^{-\lambda-1}\,\mathbb{P}[X,Y|B^t,\Theta^t]\,.$$

Summing over all $Y \in \{1, \dots, n\}^m$ yields

$$1 \;=\; \sum_{Y\in\{1,\dots,n\}^m} e^{-\lambda-1}\,\mathbb{P}[X,Y|B^t,\Theta^t]\,,$$

or equivalently,

$$\lambda = -1 + \log \sum_{Y\in\{1,\dots,n\}^m} \mathbb{P}[X,Y|B^t,\Theta^t]\,.$$

[1] Jensen's inequality for a concave function ϕ, numbers $x_1, \dots, x_n$ in Dom ϕ and positive weights $a_1, \dots, a_n$ states that

$$\frac{\sum_i a_i\phi(x_i)}{\sum_i a_i} \le \phi\left(\frac{\sum_i a_i x_i}{\sum a_i}\right)$$

Substitution into (8.8) leads to

$$0 = -\log \sum_{Y' \in \{1,\dots,n\}^m} \mathbb{P}[X, Y' | B^t, \Theta^t] \; + \; \log \mathbb{P}[X, Y | B^t, \Theta^t] \; - \; \log f^t(Y)\,.$$

Finally, solving for $f^t(Y)$,

$$f^t(Y) \;=\; \frac{\mathbb{P}[X, Y | B^t, \Theta^t]}{\sum_{Y' \in \{1,\dots,n\}^m} \mathbb{P}[X, Y' | B^t, \Theta^t]} \;=\; \frac{\mathbb{P}[X, Y | B^t, \Theta^t]}{\mathbb{P}[X | B^t, \Theta^t]} \;=\; \mathbb{P}[Y | X, B^t, \Theta^t]\,. \qquad (8.9)$$

This probability distribution maximizes $Q^f(B^t, \Theta^t)$ with respect to f. Furthermore, (8.7) holds with equality in the current guess (B^t, Θ^t) :

$$\begin{aligned} Q^{f^t}(B^t, \Theta^t) \;&=\; \sum_{Y \in \{1,\dots,n\}^m} \mathbb{P}[Y | X, B^t, \Theta^t] \; \log \frac{\mathbb{P}[X, Y | B^t, \Theta^t]}{\mathbb{P}[Y | X, B^t, \Theta^t]} \\ &=\; \sum_{Y \in \{1,\dots,n\}^m} \mathbb{P}[Y | X, B^t, \Theta^t] \; \log \mathbb{P}[X | B^t, \Theta^t] \;=\; \log \mathbb{P}[X | B^t, \Theta^t]\,. \end{aligned}$$

This makes sure that the likelihood in each consecutive step of the algorithm is at least as large as the likelihood of the previous parameter estimate.

We proceed by maximizing $Q^{f^t}(B, \Theta)$ with respect to (B, Θ). Using the definition of conditional probability we can rewrite the lower bound as

$$Q^{f^t}(B, \Theta) \;=\; \sum_{Y \in \{1,\dots,n\}^m} f^t(Y) \; \log \mathbb{P}[X | Y, B, \Theta] \;=\; \mathbb{E}^{f^t(Y)}\left[\, \log \mathbb{P}[X | Y, B, \Theta] \,\right] \qquad (8.10)$$

This explains the name 'Expectation Maximization'. Using (8.9) and (8.10):

$$\begin{aligned} Q^{f^t}(B, \Theta) \;&=\; \sum_{Y \in \{1,\dots,n\}^m} \mathbb{P}[Y | X, B^t, \Theta^t] \; \log \mathbb{P}[X | Y, B, \Theta] \\ &=\; \sum_{Y \in \{1,\dots,n\}^m} \mathbb{P}[Y | X, B^t, \Theta^t] \; \sum_{j=1}^{m} \log\left(\beta_{y_j}\, p_{y_j}(\, x_j \,|\, \theta_{y_j})\right) \\ &=\; \sum_{j=1}^{m} \sum_{Y \in \{1,\dots,n\}^m} \mathbb{P}[Y | X, B^t, \Theta^t] \; \log\left(\beta_{y_j}\, p_{y_j}(\, x_j \,|\, \theta_{y_j})\right) \\ &=\; \sum_{j=1}^{m} \sum_{l=1}^{n} \log\left(\beta_l\, p_l(\, x_j \,|\, \theta_l)\right) \sum_{\substack{Y \,\in\, \{1,\,\dots\,,\,n\}^m \\ y_j \,=\, l}} \mathbb{P}[Y | X, B^t, \Theta^t] \\ &=\; \sum_{j=1}^{m} \sum_{l=1}^{n} \log\left(\beta_l\, p_l(\, x_j \,|\, \theta_l)\right) \; \mathbb{P}[y_j = l | x_j, \beta_j^t, \theta_j^t] \\ &=\; \sum_{j=1}^{m} \sum_{l=1}^{n} \left\{ \log\left(\beta_l\right) \;+\; \log\left(p_l(\, x_j \,|\, \theta_l)\right) \right\} \mathbb{P}[y_j = l | x_j, \beta_j^t, \theta_j^t] \qquad (8.11) \end{aligned}$$

where

$$\mathbb{P}[y_j = l | x_j, \beta_j^t, \theta_j^t] = \frac{\beta_l^t p_l(x_j | \theta_l^t)}{\sum_{k=1}^{m} \beta_k^t p_k(x_j | \theta_k^t)}\,.$$

The lower bound (8.11) has split into two parts, the first of which only contains the unknowns B and the second involving only Θ. The maximization procedure of the latter term depends on the component density functions, so in general one has to rely on numerical methods. To maximize the first term we again use a Lagrange multiplier λ' to account for $\sum_{i=1}^{n} \beta_i = 1$:

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \beta_l} \left[\lambda' \left(\sum_{i=1}^{n} \beta_i - 1 \right) + \sum_{j=1}^{m} \sum_{l=1}^{n} \log(\beta_l) \; \mathbb{P}[y_j = l | x_j, \beta_j^t, \theta_j^t] \right] \\
&= \lambda' + \sum_{i=1}^{n} \frac{1}{\beta_l} \mathbb{P}[y_j = l | x_j, \beta_j^t, \theta_j^t] \, .
\end{aligned}
$$

After summing both sides over l, solving for λ' and substituting this value back one obtains

$$
\begin{aligned}
\beta_l &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}[y_j = l | x_j, \beta_j^t, \theta_j^t] \\
&= \frac{1}{n} \frac{\sum_{i=1}^{n} \beta_l^t p_l(x_j | \theta_l^t)}{\sum_{k=1}^{m} \beta_k^t p_k(x_j | \theta_k^t)} \, .
\end{aligned}
$$

This concludes the optimization of $Q^{f^t}(B, \Theta)$ with respect to (B, Θ). The algorithm is illustrated graphically in Figure 8.3.

By construction, the likelihood of the sequence of EM estimates increases monotonically. This sequence thus has a (possibly infinite) limit $\ell^* := \ell(B^*, \Theta^* | X)$. Redner and Walker [19] show (Theorem 4.1, p. 218) that ℓ^* is finite if the likelihood function ℓ and the lower bound Q are continuous. If furthermore ℓ and Q are differentiable in (B^*, Θ^*), then the derivative of the likelihood function to each of the parameters is zero in this point, a necessary condition for a maximum likelihood estimate.

The speed of convergence generally is slow: it can be shown that under certain conditions [19, Theorem 5.2] local convergence is linear.

The main advantage of EM over alternative methods such as (quasi-) Newton and Conjugate Gradient methods, some of which exhibit superlinear convergence, is its good global convergence properties.

To demonstrate the algorithm, we generate a sample from a Normal–Gumbel mixture via the methods described in Section 7 and try to retrieve the parameters by calibrating a Normal–Gumbel copula via the EM algorithm. Figure 8.4 shows the relative estimation error (i.e. the average relative error of the parameters specifying the copula) as a function of the number of steps of the algorithm. The line flattens, but does not go to zero. This means that the algorithm converges but not to the parameters that were used to generate the sample. To achieve the latter we also have to increase the sample size. Indeed, Figure 8.5 shows that the relative error tends to zero when the sample size grows. For the error to be less than 5% approximately 10000 observations are needed in this example.

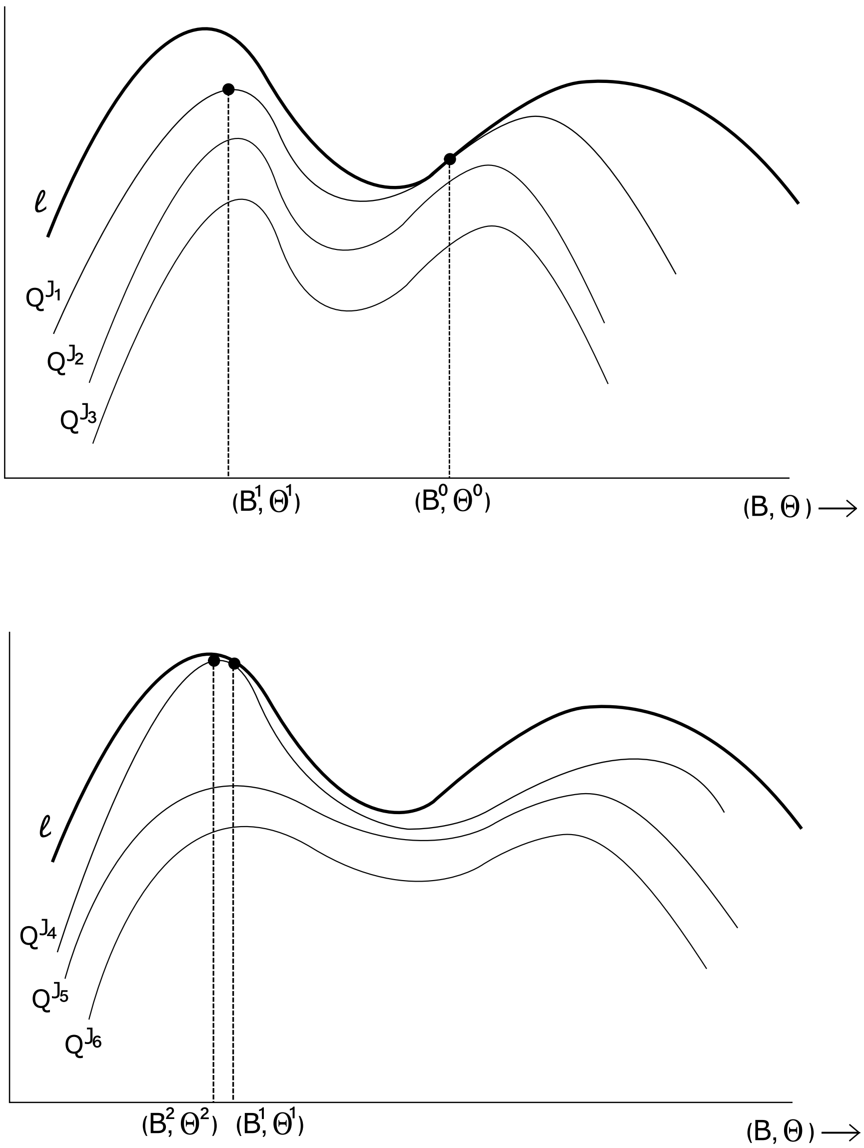


Figure 8.3: *Two steps of the EM algorithm. In each step the algorithm finds the distribution J that optimizes the lower bound Q to the likelihood function ℓ in the current estimate. Fixing the distribution, the lower bound is then maximized with respect to (B, Θ). This leads to the next estimate.*

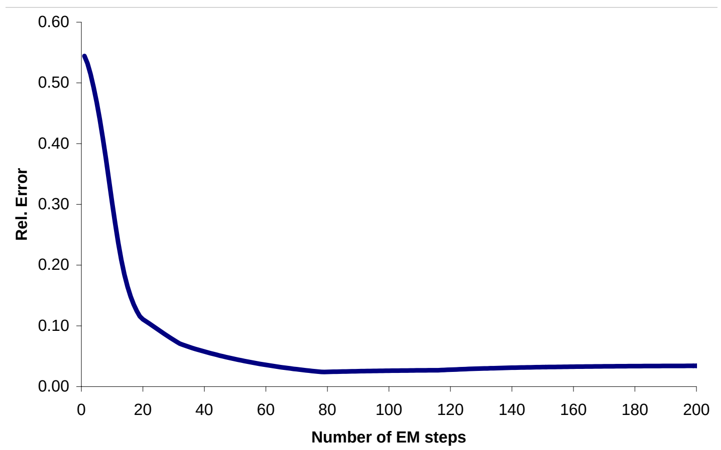


Figure 8.4: *Estimation error in consecutive steps of the EM algorithm. Calibration to Normal–Clayton mixture.*

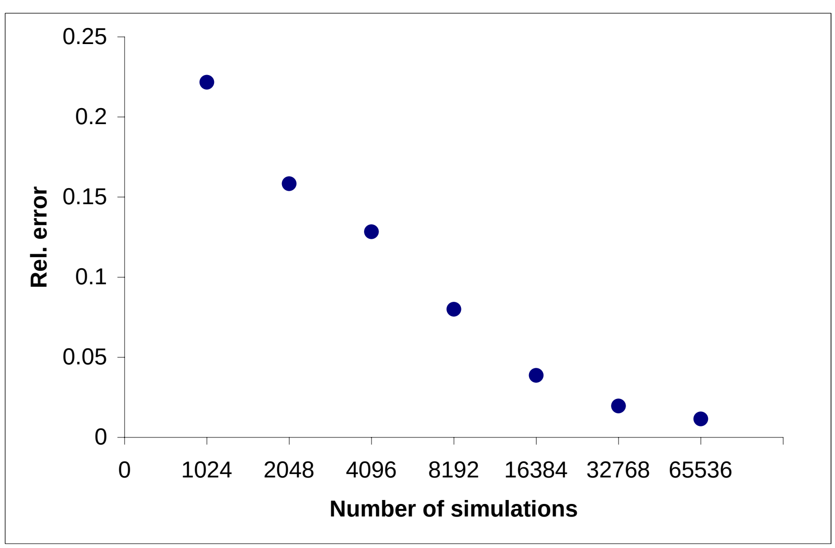


Figure 8.5: *Error of the EM algorithm for different sample sizes. Calibration to Normal–Clayton mixture.*

9 Results

We will now apply the techniques introduced in the previous sections to historical market data. First, based on Section 8, the impact of incorporating tail dependence on the quality of fitted copulas is discussed (Section 9.1). Next, using the pricing model set out in Section 6, we will assess the implication of changing copulas on derivatives prices (Section 9.2). Finally, hedging performance is compared for a model that uses a Gaussian dependence structure and one that takes into account tail dependence (Section 9.3).

9.1 Calibration results

We calibrate different copulas to historical data, compare their quality of fit and discuss whether the tail dependence in the calibrated copulas coincides with what we expect from Figures 1.1–1.6. We study the dependence between daily log-returns of the underlyings listed in Section 1.2.

As a measure for the quality of fit we will look at the likelihood (Section 8.1) of observing the historical data given some copula. This likelihood is optimized in the calibration.

An alternative measure we will consider is the L^2-norm of the difference between the estimated copula C and the empirical copula C^{emp}, that is,

$$|| \, C - C^{\text{emp}} \, ||_{L^2}^2 \quad = \quad \iint_{(u,v)\in(0,1)^2} | \, C(u,v) - C^{\text{emp}}(u,v) \, |^2 \, du \, dv \, ,$$

where, given some sample $\{(x_i, y_i)\}_{i=1}^n$ from a continuous bivariate distribution, the empirical copula is defined as

$$C^{\text{emp}}(u,v) = \frac{\text{number of pairs } (x,y) \text{ in the sample such that } x \leq x_{(i)} \text{ and } y \leq y_{(j)}}{n} \, ,$$

in which $i = \lfloor un \rfloor$, $j = \lfloor vn \rfloor$ and $\{x_{(i)}\}_{i=1}^n$, $\{y_{(i)}\}_{i=1}^n$ denote the order statistics. This L^2-norm can only be used to compare the quality of fitted copulas and *not* for hypothesis testing because its asymptotic laws depend not only on the copula, but also on the marginal distributions of the data the copula was calibrated to [20].

In calculating the likelihood, each element in the data set is used only once. Getting a reasonable approximation of the L^2-norm between the empirical and the calibrated copula however is much more expensive. Therefore we will not use the L^2-norm for calibration, but only to assess the impact of choosing a different calibration criterion on the relative quality of fit of different copulas.

Table 9.1 shows calibration results for **Nikkei 225 and SP500**. The normal copula is immediately seen not to have the highest likelihood. Instead, the copulas with upper tail dependence (Gumbel, Clayton survival) show a better fit to the data. This indeed is consistent with Figure 1.5 which also suggested upper tail dependence. Mixtures of copulas, such as Normal and Clayton survival, calibrated via the EM algorithm described in Section 8.4 can even further improve the likelihood. Figure 9.1 shows a density plot of this copula together with the diagonal section. In these pictures the tail dependence is clearly visible.

As for the mixtures, it should be noted that their likelihood should always be at least the minimum of those of the individual copulas. Otherwise assigning zero weight to one of the

Table 9.1: *Copulas calibrated to Nikkei 225 and SP500 daily logreturns.*

Copula		**Parameter**	**Tail**		**Likelihood**	$\|\|C - C^{\mathbf{emp}}\|\|_{L^2}^2$
			Lower	**Upper**		
100.00%	Normal	0.239	0.000	0.000	14.18	0.0387
100.00%	Gumbel	1.201	0.000	0.219	21.26	0.0368
100.00%	Gumbel survival	1.144	0.168	0.000	10.0881	0.0417
100.00%	Clayton	0.201	0.032	0.000	6.39	0.0447
100.00%	Clayton survival	0.394	0.000	0.172	20.72	0.0363
100.00%	Frank	1.403	0.000	0.000	13.11	0.0382
23.62%	Normal	-0.230	0.000	0.259	22.193	0.0373
76.38%	Clayton survival	0.641				
23.74%	Gumbel	1.812	0.000	0.155	21.708	0.0361
76.26%	Clayton survival	0.211				

copulas would improve the fit. So either a mixture improves the calibration, or it is not listed because one of the copulas 'drops out'.

Calibration results for **gold and copper** are shown in Table 9.2. This pair was believed to have lower tail dependence from Figure 1.3. The likelihood of copulas having lower tail dependence (Clayton, Gumbel survival) indeed is higher than the others, and in particular higher than the likelihood of the normal copula.

In case of **corn and wheat**, single alternative copulas do not improve the likelihood of observing the historical data set — see Table 9.3. Mixtures that contain the normal copula do lead to a better fit though (Normal and Clayton, Clayton survival, Gumbel survival). Looking at the empirical copula suggests the presence of lower tail dependence (Figure 1.6) but this is not evident from the calibration.

Calibration to historical observations from **USD swap futures and SP500 futures** (Table 9.4) shows that both copulas with upper and with lower tail dependence are an improvement over the normal copula. This is in agreement to Figures 1.1 and 1.2 that suggest these underlyings to have two-sided tail dependence. The density and the diagonal section of the best fit (Clayton and Gumbel) are shown in Figure 9.2.

We thus see that tail dependence properties that are observed empirically also show up in the calibration, i.e. copulas with the right tail properties improve the fit to historical data. We now turn to the question of what effect changing copulas has on prices of derivatives.

9.2 Pricing results

This section addresses the following question: how does changing from the normal copula to one that accounts for tail dependence affect the prices of derivatives. Since the normal copula is usually calibrated by direct calculation of the correlation coefficient, we will first look at the effect of changing the copula while keeping Spearman's rank correlation coefficient constant (Section 9.2.1). Thereafter, prices will be compared in case each copula is calibrated via maximum likelihood (Section 9.2.2).

Table 9.2: *Copulas calibrated to gold and copper daily logreturns.*

Copula		Parameter	Tail Lower	Tail Upper	Likelihood	$\|\|C - C^{\mathbf{emp}}\|\|_{L^2}^2$
100.00%	Normal	0.116	0.000	0.000	12.222	0.0072
100.00%	Gumbel	1.060	0.000	0.076	8.191	0.0082
100.00%	Gumbel survival	1.077	0.097	0.000	18.237	0.0072
100.00%	Clayton	0.148	0.009	0.000	16.023	0.0075
100.00%	Clayton survival	0.092	0.000	0.001	5.868	0.0091
100.00%	Frank	0.641	0.000	0.000	10.108	0.0073

Table 9.3: *Copulas calibrated to daily logreturns of 1-year corn and wheat futures.*

Copula		Parameter	Tail Lower	Tail Upper	Likelihood	$\|\|C - C^{\mathbf{emp}}\|\|_{L^2}^2$
100.00%	Normal	0.596	0.000	0.000	402.252	0.0028
100.00%	Gumbel	1.607	0.000	0.461	374.078	0.0065
100.00%	Gumbel survival	1.595	0.456	0.000	363.123	0.0082
100.00%	Clayton	0.893	0.460	0.000	300.049	0.0170
100.00%	Clayton survival	0.930	0.000	0.475	317.767	0.0151
100.00%	Frank	4.210	0.000	0.000	363.469	0.0054
87.48%	Normal	0.652	0.006	0.000	406.096	0.0028
12.52%	Clayton	0.225				
82.62%	Normal	0.652	0.038	0.000	405.978	0.0029
16.38%	Gumbel survival	1.219				
81.50%	Normal	0.660	0.000	0.038	406.514	0.0025
18.50%	Clayton survival	0.438				

Table 9.4: *Copulas calibrated to daily logreturns of 2-year SP500 futures and 2-year USD 5-year swap futures.*

Copula		Parameter	Tail Lower	Tail Upper	Likelihood	$\|\|C - C^{\mathbf{emp}}\|\|_{L^2}^2$
100.00%	Normal	0.123	0.000	0.000	13.774	0.0052
100.00%	Gumbel	1.078	0.000	0.098	17.122	0.0050
100.00%	Gumbel survival	1.075	0.095	0.000	16.016	0.0044
72.15%	Normal	-0.039	0.000	0.115	17.621	0.0048
27.78%	Gumbel	1.540				
28.30%	Normal	0.473	0.000	0.000	19.410	0.0058
71.70%	Clayton	0.001				
80.29%	Clayton	0.071	0.071	0.017	19.942	0.0054
19.71%	Gumbel	1.428				

Figure 9.1: *Diagonal section (left) and density plot (right) of Normal – Clayton survival mixture calibrated to NIKKEI 225 and SP500 daily log-returns.*

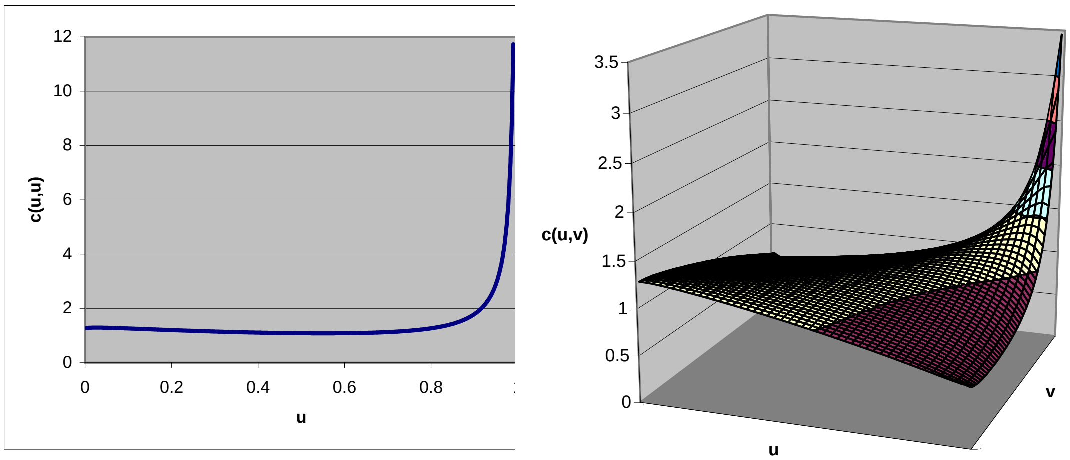
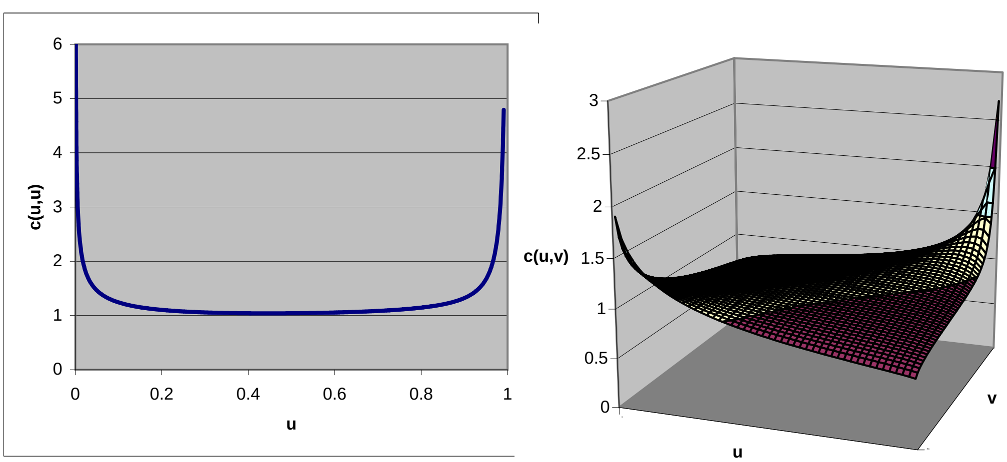


Figure 9.2: *Diagonal section (left) and density plot (right) of Clayton – Gumbel mixture calibrated to daily log-returns of 2-year futures on SP500 and USD 5-year swaps.*



9.2.1 Calibration via Spearman's rho

To understand the effect of changing copulas on option prices independently from the effect of changing the calibration procedure, we will price some of the contracts introduced in Section 1.2 — namely a best-of returns option (1.3), a worst-of returns option (1.4) and an at-the-money spread (1.6) — using different copulas while keeping Spearman's rho fixed.

For the copulas under consideration (Normal, Clayton and Gumbel) there exist one-to-one relationships between Spearman's rank correlation coefficient and the copula parameter. Note that the Gumbel copula can only capture positive dependence. Therefore we will only consider positive rank correlation values.

The prices are calculated via 65535 Monte Carlo simulations. Observations from the copula are drawn using the algorithms described in Section 7. Lognormal marginal distributions were chosen with implied volatility $\sigma = 0.3$. The results are shown in Figures 9.3, 9.4 and 9.5.

For the normal case one can prove [21] that spread prices are a decreasing function of the correlation parameter. This is indeed seen in Figure 9.3. The Clayton copula gives a higher price over the whole range of positive dependence with a maximum of 15% relative difference whereas Gumbel and Clayton survival have a negative effect on the price that does not exceed 10%.

The price of a worst-of returns contract can be shown to be positively related to correlation [21]. This is also clear from Figure 9.4. Prices for other copulas show a relative difference of less than 1%. In case of Gumbel and Clayton survival this difference is positive, for Gumbel survival and Clayton it is negative.

Finally also Figure 9.5 agrees with theoretical results in [21]: best-of returns option prices are a decreasing function of correlation. Only the Clayton copula leads to a higher price, the others give comparable or slightly lower prices. Relative differences remain within 1%.

9.2.2 Calibration via maximum likelihood

In the previous section we compared pricing results for different copulas while keeping Spearman's rank correlation coefficient fixed. We will now compare prices where each copula is calibrated via maximum likelihood using the pricing model described in Section 6.2.2. Copulas are calibrated to historical NIKKEI 225 and SP500 prices. Lognormal margins are used with historical volatility.

Tables 9.5, 9.6 and 9.7 list premia for worst-of, best-of and spread options with a 1-year maturity that are priced using different copulas. These copulas are calibrated via maximum likelihood to daily returns, monthly returns and levels. Prices for a Gumbel–Clayton survival mixture in case of calibration to levels are not given since the likelihood of the combination did not exceed the likelihood of the single copulas.

Spearman's rank correlation for the estimated copulas is seen to be approximately 0.15 for all copulas calibrated to daily returns, 0.25 in case of monthly returns and much higher for copulas calibrated to levels. The increasing rank correlation is a result of the calibration procedure assuming the time series not to be autocorrelated (Section 8.1). In reality consecutive returns become more dependent for larger time windows (see Table 6.1) leading to an overestimation

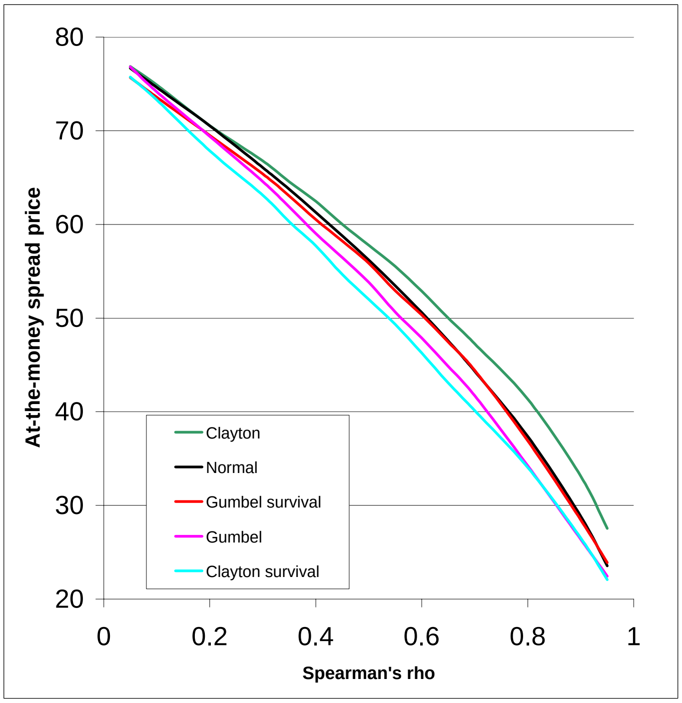


Figure 9.3: *At-the-money spread prices in terms of Spearman's rho for different copulas.*

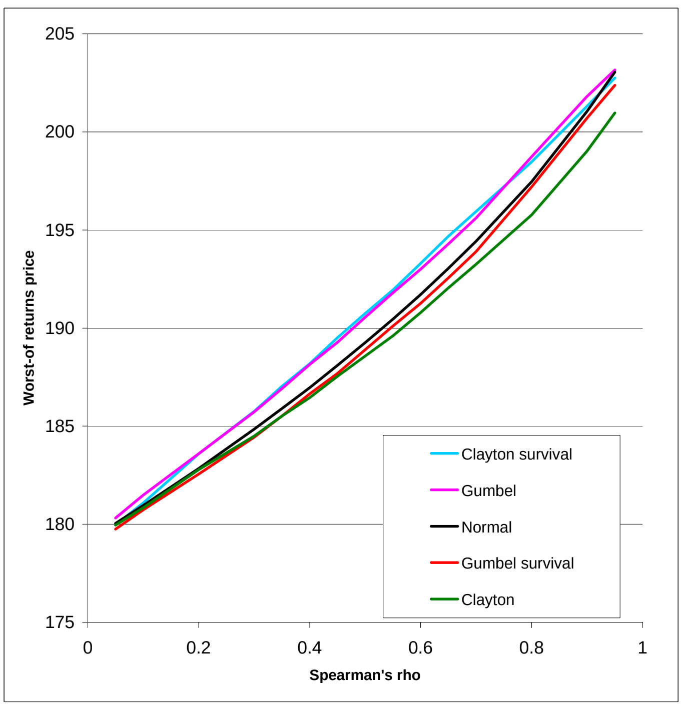


Figure 9.4: *Worst-of returns prices in terms of Spearman's rho for different copulas.*

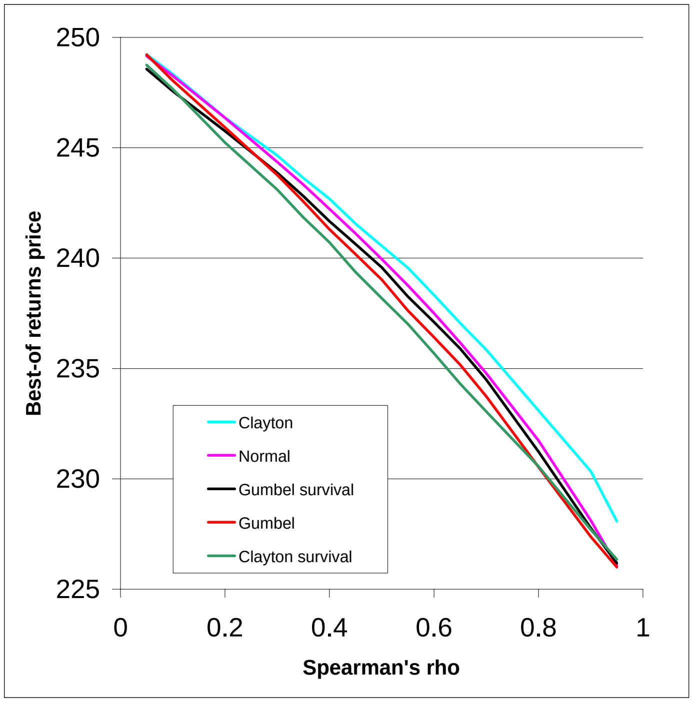


Figure 9.5: *Best-of returns prices in terms of Spearman's rho for different copulas.*

of the likelihood of copulas with high dependence [22, Figure 3]. Calibration to levels thus is, from a practical and theoretical point of view, hard to defend.

Relative price differences due to changing the copula in the multi-multi model are roughly around 1% for the contracts under consideration if one calibrates to periodic returns (either daily or monthly). When calibrating to levels outcomes can deviate substantially.

Reason for the price differences being very small in case of calibration to daily returns is the central limit theorem: the pricing model essentially calculates the convolution of the distribution of daily price changes and this convolution tends to a multivariate normal distribution. The 'shape' of the terminal dependence structure thus will be the same for all copulas.

The question remains whether these new prices are an improvement. This can only be answered by assessing the model's ability to hedge the options it is expected to price.

Table 9.5: *Worst-of returns prices for different copulas calibrated to daily returns, monthly returns and levels. Underlyings are NIKKEI 225 and SP500.*

Copula	Daily returns			Monthly returns			Levels		
	ρ_S	Price	Rel. diff.	ρ_S	Price	Rel. diff.	ρ_S	Price	Rel. diff.
Normal	0.158	5.05	0.00%	0.240	5.00	0.00%	0.916	6.97	0.00%
Gumbel	0.142	4.98	-1.39%	0.257	5.07	1.40%	0.917	7.05	1.17%
Clayton survival	0.110	4.88	-3.37%	0.260	5.12	2.40%	0.877	6.98	0.18%
Normal Clayton survival	0.156	5.04	-0.20%	0.231	5.03	0.60%	0.934	7.06	1.37%
Gumbel Clayton survival	0.140	5.00	-0.99%	0.265	5.13	2.60%	–		

Table 9.6: *Best-of returns performance prices for different copulas calibrated to daily returns, monthly returns and levels. Underlyings are NIKKEI 225 and SP500.*

Copula	Daily returns			Monthly returns			Levels		
	ρ_S	Price	Rel. diff.	ρ_S	Price	Rel. diff.	ρ_S	Price	Rel. diff.
Normal	0.158	14.74	0.00%	0.240	14.68	0.00%	0.916	12.76	0.00%
Gumbel	0.142	14.77	0.20%	0.257	14.65	-0.20%	0.917	12.71	-0.39%
Clayton survival	0.110	14.79	0.34%	0.260	14.66	-0.14%	0.877	12.74	-0.11%
Normal Clayton survival	0.156	14.76	0.14%	0.231	14.72	0.27%	0.934	12.66	-0.78%
Gumbel Clayton survival	0.140	14.78	0.27%	0.265	14.63	-0.34%	–		

Table 9.7: *At-the-money spread prices for different copulas calibrated to daily returns, monthly returns and levels. Underlyings are NIKKEI 225 and SP500.*

Copula	Daily returns			Monthly returns			Levels		
	ρ_S	Price	Rel. diff.	ρ_S	Price	Rel. diff.	ρ_S	Price	Rel. diff.
Normal	0.158	13.63	0.00%	0.240	13.77	0.00%	0.916	13.45	0.00%
Gumbel	0.142	13.62	-0.07%	0.257	13.80	0.22%	0.917	13.49	0.29%
Clayton surv.	0.110	13.67	0.29%	0.260	13.88	0.85%	0.877	13.49	0.32%
Normal Clayton surv.	0.156	13.71	0.59%	0.231	13.86	0.67%	0.934	13.45	0.01%
Gumbel Clayton surv.	0.140	13.65	0.15%	0.265	13.84	0.53%	–		

9.3 Hedging

Pricing of financial contracts is based on absence of arbitrage, i.e. the argument that an option whose payoff at maturity coincides with the value of a self-financing portfolio[2] of liquid tradable assets should have the same value as this portfolio at every prior point in time. Therefore, given a certain pricing model, it is important that the price of the replicating portfolio follows the price of the option closely. A replicating portfolio can be constructed from the pricing model by calculating the sensitivity of the contract's value to the prices of the hedging instruments and assembling the portfolio accordingly. These sensitivities are referred to as 'the greeks' because they are usually denoted by a greek letter. For our hedging we will consider the 'deltas', i.e. the first partial derivatives of the option price to the prices of the underlying assets.

Definition of delta neutrality Let V be the present value of a position in the option that we seek to delta hedge. Suppose the price of this option depends on the market observable quantities $x_1, \ldots, x_n$. These can for instance be prices of assets, prices of other contracts or interest rates. To hedge the option we take positions $\xi_1, \ldots, \xi_n$ in n hedging instruments whose present values will be denoted $H_1, \ldots, H_n$ respectively. To make the option position 'delta neutral' the $\xi_1, \ldots, \xi_n$ have to satisfy

$$-\begin{pmatrix} \frac{\partial V}{\partial x_1} \\ \vdots \\ \frac{\partial V}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial H_1}{\partial x_1} & \cdots & \frac{\partial H_n}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial H_1}{\partial x_n} & \cdots & \frac{\partial H_n}{\partial x_n} \end{pmatrix} \cdot \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix}. \tag{9.1}$$

In theory the portfolio should be rehedged continuously, meaning that the positions should be updated constantly in order to satisfy (9.1) for all points in time. In practise continuous rehedging is not possible. Instead, we will update the portfolio on a daily basis.

Hedging performance can be improved by rehedging more frequently or by considering higher order partial derivatives, such as the 'gammas', i.e. second order partial derivatives of the option price to the underlying assets.

Delta hedging of a European option on two assets We will now apply this principle for the particular case of a European option on two underlyings, such as a spread, best-of or worst-of. As hedging instruments we choose futures on the two underlyings and a zero coupon bond[3], all of which have the same maturity as the option we are hedging. These futures are quoted in the market by their 'par strike', i.e. by the strike that makes the value of the contract zero. The present value of a contract whose strike is K where the current par strike is K^{par} equals

$$\text{Present value future} = \text{Discount factor to maturity} \times \left(K^{\text{par}} - K\right). \tag{9.2}$$

The partial derivatives of the futures' present values can be calculated via (9.2). The value V of the European option at time t depends on the par strikes $F_1(t,T), F_2(t,T)$ of the underlyings

[2] A replicating portfolio is called self-financing if no money injections or withdrawals are needed to ensure its value to coincide with some option's payoff at maturity.

[3] A zero coupon bond is a contract that pays one unit currency at maturity.

and the value of the bond $B(t,T)$ with maturity T. The partial derivatives of the option value usually cannot be calculated analytically. We must therefore fall back on finite differences. In case of a bivariate European option that is hedged with futures on the underlyings and a bond, for small ε, the solution of system (9.1) is approximated by

$$\begin{aligned}
\text{Position in future 1} &= \frac{V(\,(1+\varepsilon)F_1(t,T),\, F_2(t,T),\, B(t,T)) - V(F_1(t,T),\, F_2(t,T),\, B(t,T))}{\varepsilon\, F_1(t,T)\, B(t,T)}\,, \\
\text{Position in future 2} &= \frac{V(F_1(t,T),\, (1+\varepsilon)\, F_2(t,T),\, B(t,T)) - V(F_1(t,T),\, F_2(t,T),\, B(t,T))}{\varepsilon\, F_2(t,T)\, B(t,T)}\,, \\
\text{Position in bond} &= V(F_1(t,T),\, F_2(t,T),\, B(t,T))\,.
\end{aligned}$$

Payoff formulas frequently are not differentiable in all points of their domain, for example when the payoff is 'capped' or 'floored'. The greeks of these contracts then are discontinuous functions of their inputs. If Monte Carlo methods are used, these discontinuities may require a larger number of simulations in order for the greeks to converge.

Hedge test To assess whether a certain model is able to succesfully hedge the contracts that are priced with it, we construct and maintain a delta-neutral portfolio and investigate its fluctuations in value:

- The option is sold and the premium is put in a money account earning the overnight rate.
- The portfolio is delta hedged using futures on the underlying assets and zero coupon bonds.
- The portfolio is revalued and rebalanced in the same way on each day of the simulation period. Every day the hedging instruments are liquidated and replaced to re-establish delta-neutrality.

Ideally the value of the hedged portfolio would continuously have zero mean and variance. In practise this will not be the case since we are hedging in descrete time. However, we still aim for these numbers to be as small as possible in absolute value.

To be able to compare hedging performance for different products, the mean and variance are scaled by the (initial) value of the option position. Intuitively one should also consider the fact that the prices of certain contracts, by nature of their underlyings, are more volatile than others. This suggests looking at the quotient of the standard deviation of the values of the hedged portfolio and the standard deviation of the prices of the item being hedged. The 'volatility reduction' will be defined as:

$$\text{Volatility reduction} = 1 - \frac{\text{Standard deviation hedged portfolio}}{\text{Standard deviation naked option position}}\,.$$

Market incompleteness issues The risk involved in the (stochastic) discount factor of the payoff's expected value is delta hedged using a bond with the same maturity as the contract that is being hedged. In practise these bonds might not always be available in the market. In these cases a 'virtual' bond is constructed by linear interpolation between zero rates obtained from deposit and swap rates.

Table 9.8: *Hedging performance of the multi-multi model with a normal copula (no tail dependence) and a Normal–Clayton survival mixture copula (upper tail dependence).*

	Average		**Standard deviation**		**Volatility**
	Option Pos.	Hedging ptf.	Option Pos.	Hedging ptf.	**reduction**
Call Return Corn	1.514	-0.044	-0.700	-0.016	0.977
Best-of Returns					
Normal copula	1.455	-0.021	-0.333	-0.009	0.973
Normal–Clayton surv.	1.985	0.021	-1.028	-0.011	**0.990**
Worst-of Returns					
Normal copula	1.760	-0.005	-0.880	-0.009	0.989
Normal–Clayton surv.	1.985	0.021	-1.028	-0.011	**0.990**
Spread Returns					
Normal copula	1.253	-0.039	-0.338	-0.015	0.957
Normal–Clayton surv.	1.449	-0.008	-0.441	-0.006	**0.987**
ATM Spread					
Normal copula	1.311	-0.168	-0.658	-0.070	**0.894**
Normal–Clayton surv.	1.860	0.217	-1.239	-0.236	0.809

Effect of tail dependence on hedging To get an impression of the hedging performance of European options on tail dependent underlyings, we perform a hedge test for different options on corn and wheat. In this test we use futures prices of corn and wheat taken from the Chicago Board of Trading (CBOT) over a one-year period from December 2005 until December 2006. The marginal terminal distributions are modelled lognormal where the volatility is set to the in-sample historical volatility. We will use two models for calculating the greeks. The first one is the multi-multi model with a normal copula calibrated as described in Section 6.2.1. The second is multi-multi with a 11% Normal ($\rho = 0.78$) – 89% Clayton survival ($\theta = 2.45$) mixture copula that was calibrated to historical levels.

We will compare the volatility reduction of a vanilla call on return option

$$\text{Call on return} = \max\left(0, \frac{S(T)}{S(0)} - 1\right),$$

to the hedging performance of a best-of returns option, a worst of returns option, a spread on returns and an at-the-money spread option — see equations (1.3)–(1.6).

The results are summarized in Table 9.8. The hedging portfolio should replicate the option as closely as possible, that is, the average and the variance of the hedged portfolio have to be close to zero. In case of the normal copula, volatility reduction for the multi-asset contracts is of the same order of magnitude as for the vanilla call option, except for the at-the-money spread option whose hedging performance is slightly worse. This is also seen in Figures 9.6–9.9 which show the value of the hedged portfolio as it changes over time. This hedge test thus suggests that not taking into account tail dependence in pricing best-of, worst-of and spread options on returns does not lead to hedging performance that is significantly worse than for a vanilla call option.

The mixture copula model that incorporates upper tail dependence does not significantly improve volatility reduction for the multi-asset options. For the at-the-money spread option, the normal copula outperformes the normal-Clayton survival mixture.

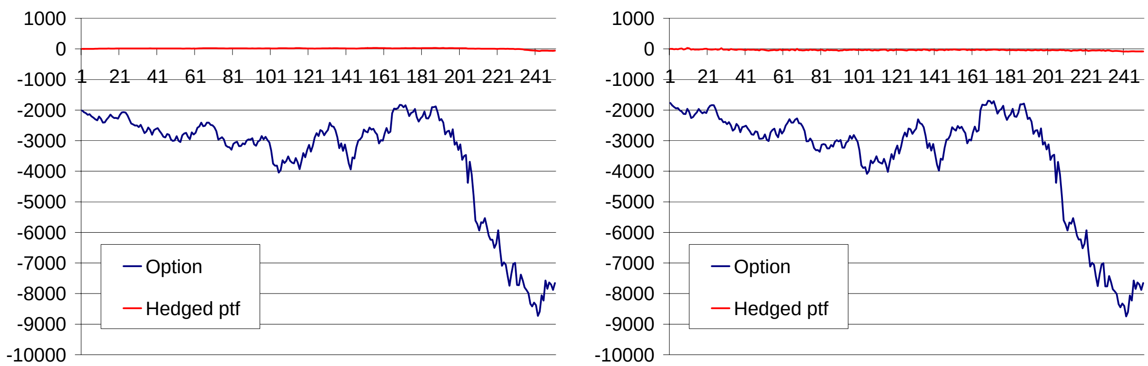


Figure 9.6: *Daily profit and loss for a worst-of returns option hedged using a normal copula (left) and using a normal-Clayton mixture copula with upper tail dependence (right).*

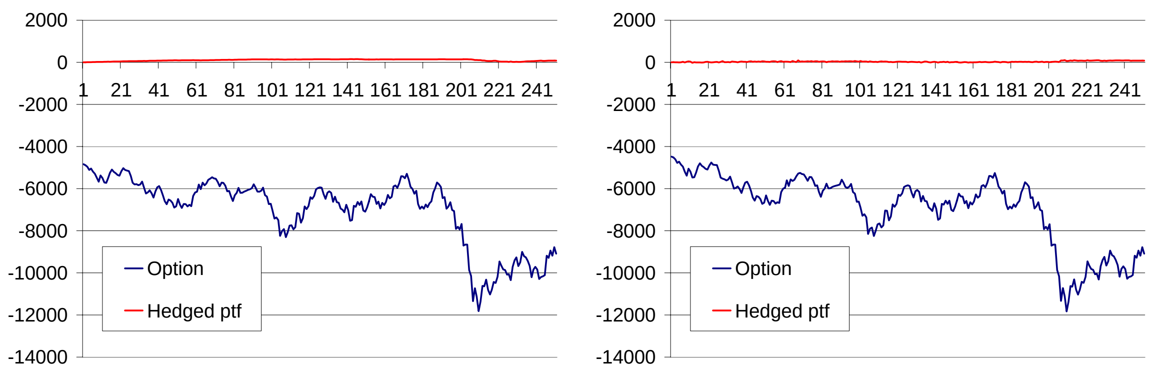


Figure 9.7: *Daily profit and loss for a best-of returns option hedged using a normal copula (left) and using a normal-Clayton mixture copula with upper tail dependence (right).*

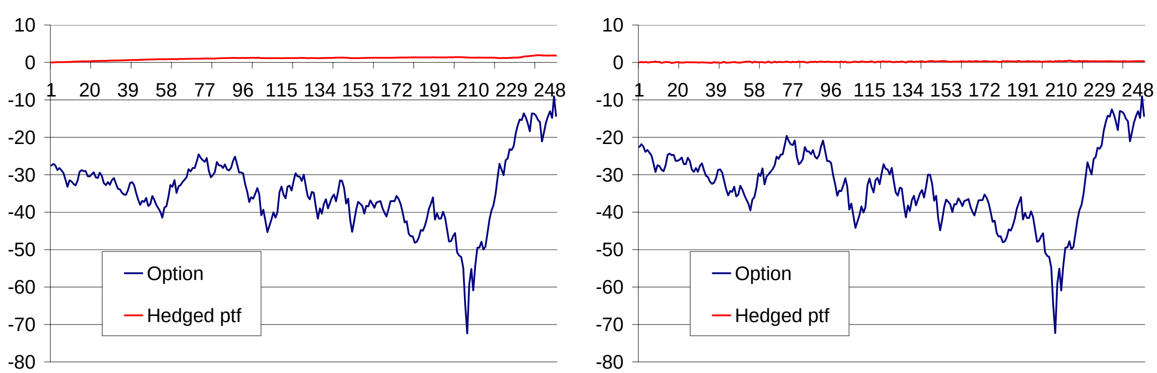


Figure 9.8: *Daily profit and loss for a spread on returns option hedged using a normal copula (left) and using a normal-Clayton mixture copula with upper tail dependence (right).*

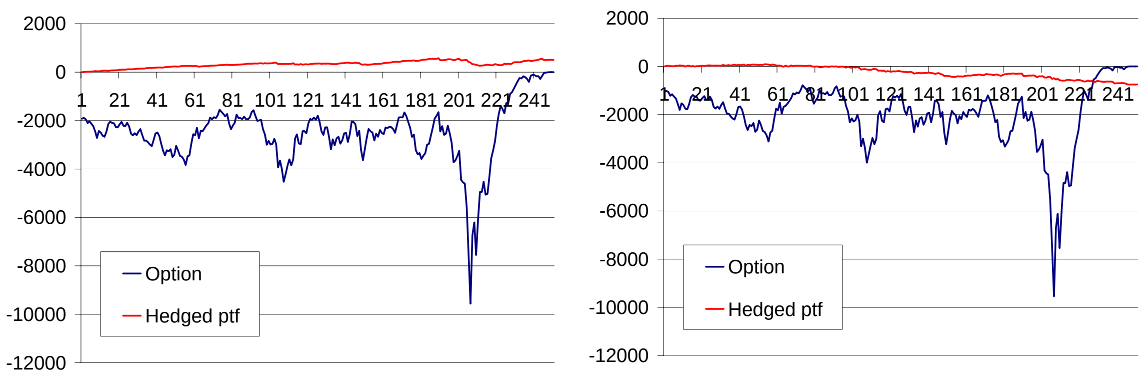


Figure 9.9: *Daily profit and loss for an at-the-money spread option hedged using a normal copula (left) and using a normal-Clayton mixture copula with upper tail dependence (right).*

10 Conclusions and recommendations

The multivariate Black-Scholes pricing model uses a Gaussian copula to model the dependence between underlyings. However, empirical tests show that this copula underestimates the probability of extreme events. For some financial contracts which depend on the tails of the terminal distributions of the underlyings — such as worst-of, best-of and spread options, see equations (1.3)–(1.6) — this may have implications on the price.

Empirical evidence of tail dependence is most profound in daily returns, much less in the levels of the underlyings (Table 1.1). This means that extreme jumps in daily price changes are possibly underestimated by a Gaussian copula, extreme 'overall' levels of the stock are not.

The concept of copulas is very appealing since we can replace the Gaussian copula while leaving the marginal distributions intact to be consistent with models for single-asset derivatives. The new copula needs to have the right amount of upper and lower tail dependence. Most of the widely-used copulas are parametrized in a way that upper and lower tail dependence cannot be controlled independently. Therefore, Hu [22] suggests to calibrate mixtures — i.e. affine linear combinations — of copulas to historical stock returns.

In case maximum likelihood is chosen as the calibration criterion, an optimization problem has to be solved numerically. A common approach for mixture distributions is to use the Expectation Maximization algorithm (Section 8.4). Calibration to historical returns (Tables 9.1–9.4) leads to copula estimates whose tail properties coincide with empirical observations. Also, these copulas show a better fit to the data according to the distance between the empirical and the calibrated copula.

If a Gaussian copula is used the covariance matrix of the terminal joint distribution can be taken from the multivariate Black-Scholes model even if the marginal distributions are not lognormal. For other copulas, that are not parametrized in terms of covariances, this is not possible. Therefore we will have to numerically approximate the convolution of the distribution of daily or monthly price changes (Section 6.2.2). This slows down computation and requires a larger Monte Carlo simulation.

As for the impact on pricing, we have to distinguish two effects: the first is the change of copula and the second is the change of the calibration procedure since in traditional models calibration of the dependence structure is based on correlation rather than maximum likelihood.

To single out the effect of changing the copula, we first fix Spearman's rank correlation coefficient. The shift of prices due to changing copulas for worst-of and best-of returns contracts remain within 1% relative difference (Figures 9.5 and 9.4). For at-the-money spread options the relative differences are larger (Figure 9.3).

Calibration to daily and monthly returns via maximum likelihood leads to relative price differences around 1% (Tables 9.5–9.7). Spearman's rank correlation hardly varies per calibrated copula, but increases with the period of the returns (i.e. daily or monthly). This is due to autocorrelation being more profound for larger return windows (Table 6.1) contradicting the maximum likelihood assumption of independent observations.

For the Expectation Maximization algorithm to converge, a large number of observations is needed (Figure 8.5). When calibrating to non-overlapping monthly returns, the necessary amount of data is hard to attain.

Calibration to levels leads to prices that are much different from the ones that stem from calibration to returns. These prices are likely to be wrong since autocorrelation in levels tends to be very high.

Hedge tests for best-of, worst-of and spread options on tail dependent underlyings (Section 9.3) show that the new model incorporating tail dependence does not clearly improve the volatility reduction (Table 9.8) as compared to the normal copula. However, in absolute sense, hedging performance of both models is not significantly worse for best-of, worst-of and spread structures than for vanilla call options.

Summarizing, the impact of taking into account tail dependence in pricing at-the-money bivariate best-of, worst-of and spread contracts is small. This is partly due to the central limit theorem leading to a Gaussian terminal dependence structure even when the returns are modelled to have heavy tails and partly due to the low strike not emphasizing the bivariate tails. For the contracts studied, hedge tests show no clear indication of problems arising from using a model that does not incorporate tail dependence in spite of the underlyings showing clear empirical evidence of being tail dependent.

Further research may be carried out to find whether the same conclusions hold when volatility smile is accounted for in the hedge tests. In this study we only considered hedging with lognormal marginal distributions.

The effect of changing copulas on prices of worst-of, best-of and spread contracts may be small, for more path dependent products tail dependence can be of greater influence since extreme jumps in daily price changes occur more frequently than extreme levels of the underlyings. To model path dependent derivatives, also dependence in the time direction has to be taken into account. This is more difficult since it is not clear how this can be done while still being consistent with the time dependence contained in the processes that model the forward prices of the underlyings. Also this would require using higher dimensional copulas with mutually distinct bivariate margins (Section 5.4).

A Basics of derivatives pricing

Consider a market model consisting of price processes

$$S(t) = \begin{pmatrix} S_1(t) \\ \vdots \\ \vdots \\ S_n(t) \end{pmatrix} \tag{A.1}$$

defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Also, let $B(t)$ denote the value of the money account at time t. For the time being, assume the interest rate to be deterministic.

Let $\mathcal{F}_t^S$ denote the sigma algebra generated by S over the interval $[0, t]$:

$$\mathcal{F}_t^S = \sigma\{(S_1(s), \ldots, S_n(s)) : s \leq t\}.$$

Intuitively, an event belongs to the sigma algebra generated by S over $[0, t]$ if, from the trajectory of S(t) over $[0, t]$, it is possible to decide whether the event has occured or not.

A **T**-claim is any $\mathcal{F}_{\mathbf{T}}^S$-measurable random variable $\mathcal{X}$.

Question: <u>what should be the price $\Pi(t; \mathcal{X})$ of the T-claim $\mathcal{X}$ at time t?</u>

A.1 No arbitrage and the market price of risk

To be able to assign a price to a derivative, the market is assumed to be **arbitrage free**, i.e. it is not possible to make a risk-free profit. The next characterisation of risk-free markets will be used extensively throughout this section.

Consider two assets S_1, S_2 driven by the same Wiener process, i.e.

$$\begin{aligned} dS_1 &= \mu_1 \, dt + \sigma_1 \, dW, \\ dS_2 &= \mu_2 \, dt + \sigma_2 \, dW, \end{aligned}$$

but with different volatilities $\sigma_1 \neq \sigma_2$. Construct a portfolio V based on S_1 and S_2 as follows:

$$V = \frac{\sigma_2}{\sigma_2 - \sigma_1} S_1 + \frac{-\sigma_1}{\sigma_2 - \sigma_1} S_2.$$

This combination eliminates the dW-term from the V-dynamics:

$$dV = \left[\frac{\sigma_2}{\sigma_2 - \sigma_1} \mu_1 + \frac{-\sigma_1}{\sigma_2 - \sigma_1} \mu_2 \right] dt.$$

Thus, the portfolio is risk-free. The no arbitrage assumption requires

$$\frac{\sigma_2}{\sigma_2 - \sigma_1} \mu_1 + \frac{-\sigma_1}{\sigma_2 - \sigma_1} \mu_2 = r,$$

where r is the interest rate. This relation is equivalent to

$$\frac{\mu_1 - r}{\sigma_1} = \frac{\mu_2 - r}{\sigma_2} := \lambda.$$

The no arbitrage condition thus entails the **market price of risk** λ to be **equal for all assets** in a market that are **driven by the same Wiener process**. This characterisation will be used in section A.3 to derive the Black-Scholes fundamental PDE.

Note that for the above argument to be valid, the assets have to be tradable and the market must be liquid, i.e. assets can be bought and sold quickly. Furthermore, it must to be possible to sell a borrowed stock (short selling). It is also assumed that there are no transaction costs, no taxes and no storage costs.

A.2 Itō formula

Theorem A.1 (Itō's formula for two standard processes) *Let f be an $\mathcal{R} \times \mathcal{R} \to \mathcal{R}$ function such that all derivatives up to order 2 exist and are square integrable. Assume the processes $X(t)$ and $Y(t)$ to follow the dynamics*

$$
\begin{aligned}
dX(t) &= a(t)\,dt + b(t)\,dW(t), \\
dY(t) &= \alpha(t)\,dt + \beta(t)\,dW(t).
\end{aligned}
$$

If $Z(t) = f(\,X(t),\, Y(t)\,)$, then

$$
\begin{aligned}
dZ(t) &= f_x(\,X(t),\, Y(t)\,)\,dX(t) + f_y(\,X(t),\, Y(t)\,)\,dY(t) \\
&\quad + \frac{1}{2} f_{xx}(\,X(t),\, Y(t)\,)\,dX(t)\,dX(t) + \frac{1}{2} f_{yy}(\,X(t),\, Y(t)\,)\,dY(t)\,dY(t) \\
&\quad + f_{xy}(\,X(t),\, Y(t)\,)\,dX(t)\,dY(t).
\end{aligned}
$$

A proof can be found in Steele [23]. Particularly useful is the case when $f(x,y) = x/y$:

Corollary A.2 (Itō's division rule) *Assume the processes $X(t)$ and $Y(t)$ to follow the dynamics*

$$
\begin{aligned}
dX(t) &= \mu_X\, X(t)\,dt + \sigma_X X(t)\,dW(t), \\
dY(t) &= \mu_Y\, Y(t)\,dt + \sigma_Y Y(t)\,dW(t).
\end{aligned}
$$

Then $Z(t) = X(t)/Y(t)$ has dynamics

$$
\begin{aligned}
dZ(t) &= \mu_Z\, Z(t)\,dt + \sigma_Z\, Z(t)\,dW(t), \\
\sigma_Z &= \sigma_X - \sigma_Y, \\
\mu_Z &= \mu_X - \mu_Y + \sigma_Y(\sigma_Y - \sigma_X).
\end{aligned}
$$

A.3 Fundamental PDE, Black-Scholes

The price of a contingent claim can be recovered by solving the fundamental PDE associated with the model.

As an example, consider the Black-Scholes model consisting of two assets with the following dynamics:

$$\begin{aligned} dB(t) &= rB(t)dt, \\ dS(t) &= \mu S(t)dt + \sigma S(t)dW(t). \end{aligned} \tag{A.2}$$

The interest rate r and the volatility σ are assumed to be constant.

The claim $\mathcal{X} = \Psi(S(T))$ has price process

$$\Pi(t) = F(t, S(t)) \tag{A.3}$$

where F is a smooth function. Applying Itō's formula to (A.3) and omitting arguments:

$$\begin{aligned} d\Pi &= \mu_\Pi \, \Pi \, dt + \sigma_\Pi \, \Pi \, dW, \\ \mu_\Pi &= \frac{F_t + \mu S F_S + \frac{1}{2}\sigma^2 S^2 F_{SS}}{F}, \\ \sigma_\Pi &= \frac{\sigma S F_S}{F}. \end{aligned}$$

No arbitrage implies the market price of risk to be the same for all assets driven by the same Wiener process:

$$\frac{\mu - r}{\sigma} = \frac{\mu_\Pi - r}{\sigma_\Pi} = \lambda,$$

so

$$\frac{F_t + (r + \lambda\sigma) S F_S + \frac{1}{2}\sigma^2 S^2 F_{SS}}{F} = \mu_\Pi = r + \lambda \sigma_\Pi = r + \lambda \frac{\sigma S F_S}{F}.$$

This yields, after rearranging terms, the fundamental PDE for the Black-Scholes model:

$$\begin{cases} F_t + rSF_S + \frac{1}{2}\sigma^2 S^2 F_{SS} = rF, \\ F(T, S(T)) = \Psi(S(T)). \end{cases}$$

For a call option, i.e. $\Psi(S(T)) = \max(0, S(T) - K)$, the solution of this PDE is the well-known Black-Scholes equation:

$$F(t, S(t)) = S(0)\Phi(d_1) - Ke^{-rT}\Phi(d_2), \tag{A.4}$$

where

$$\begin{aligned} d_1 &= \frac{\log \frac{S(0)}{K} + (r + \frac{\sigma^2}{2})T}{\sigma\sqrt{T}}, \\ d_2 &= d_1 - \sigma\sqrt{T}. \end{aligned}$$

A.4 Martingale approach

An alternative way to determine the price of a contingent claim is to exploit martingale properties. The martingale approach consists in changing the measure of the Wiener process driving the asset prices, such that, under the new measure, all tradable assets (including the money account) have the same instantaneous rate of return. From Itō's division rule it then follows that choosing the money account as the numéraire yields a process with zero drift.

Modulo a technicality, this means that each quotient of an asset price and the money account is a martingale. This leads to pricing formula (A.6). We will now repeat this argument in more detail.

First, we need to relate a change in the drift of a Wiener process to a change of measure. This relation is described by Girsanov's theorem.

Theorem A.3 (Girsanov Theorem) *Let $W^{\mathbb{P}}$ be a standard $\mathbb{P}$-Wiener process on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\phi(t)$ be a vector process that, for every t, is measurable by the sigma-algebra generated by $W^{\mathbb{P}}$ on [0,t]. If $\phi(t)$ satisfies the Novikov condition*

$$\mathbb{E}^{\mathbb{P}}\left[e^{\frac{1}{2}\int_0^T ||\phi(t)||^2\,dt}\right] < \infty, \tag{A.5}$$

then there exists a measure $\mathbb{Q}$, equivalent to $\mathbb{P}$, such that

$$\begin{aligned} \frac{d\mathbb{Q}}{d\mathbb{P}} &= e^{-\int_0^T \phi(t)\,dW^{\mathbb{P}}(t) - \frac{1}{2}\int_0^T \phi^2(t)\,dt}, \\ dW^{\mathbb{P}}(t) &= \phi(t)dt + dW^{\mathbb{Q}}(t). \end{aligned}$$

For a proof, refer to Björk [24].

How does the no arbitrage condition come across in the marginale approach? Consider an asset S with dynamics

$$dS(t) = \mu S(t)dt + \sigma S(t)dW^{\mathbb{P}}(t).$$

Then,

$$\mathbb{E}^{\mathbb{P}}\left[\frac{dS(t)}{S(t)}\right] = \mu dt := (r + \lambda\sigma)dt,$$

where λ is the market price of risk.

$$\begin{aligned} dS(t) &= (r + \lambda\sigma)S(t)dt + \sigma S(t)dW^{\mathbb{P}} \\ &= rS(t)dt + \sigma S(t)(\lambda dt + dW^{\mathbb{P}}) \end{aligned}$$

Girsanov's theorem implies the existence of a new measure $\mathbb{Q}$ such that $\lambda dt + dW^{\mathbb{P}}$ is a Wiener process:

$$dS(t) = rS(t)dt + \sigma S(t)dW^{\mathbb{Q}}$$

Under the new measure, the instantaneous rate of return on the asset equals r:

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{dS(t)}{S(t)}\right] = rdt.$$

Note that **the risk-neutral measure $\mathbb{Q}$ only depends on the market price of risk λ, which is the same for all assets in the market**. Thus, under $\mathbb{Q}$, all assets in the market have instantaneous rate of return equal to the instantaneous yield r of the risk free asset B.

From Itō's division rule (Corollary A.2) it follows that the process $S(t)/B(t)$ has zero drift under the new measure $\mathbb{Q}$. If the volatility of this process satifies the Novikov condition (A.5), then zero drift implies $S(t)/B(t)$ to be a martingale[4]. Pricing formula (A.6) is an immediate

[4] In general, zero drift does not imply a stochastic process to be a martingale. The implication holds under an extra condition, see [25] p. 79. For exponential martingales, this condition is equivalent to the (more practical) Novikov condition.

consequence of this. In the following we will denote $\mathbb{Q} =: \mathcal{M}(B(\cdot))$ to emphasize the relation between measure and numéraire.

A measure like $\mathcal{M}(B(\cdot))$ under which the prices of all assets in the market discounted by the risk-neutral bond, are martingales, is called an **equivalent martingale measure**. 'Equivalent' means that $\mathbb{P}$ and $\mathcal{M}(B(\cdot))$ agree on the same zero sets.

Theorem A.4 (First Fundamental Pricing Theorem) *If a market model has a risk-neutral probability measure, then it does not admit arbitrage.*

Theorem A.5 (General pricing formula) *The arbitrage free price process for the T-claim $\mathcal{X}$ is given by*

$$\frac{\Pi(t;\mathcal{X})}{B(t)} = \mathbb{E}^{\mathcal{M}(B(\cdot))}\left[\frac{\Pi(T;\mathcal{X})}{B(T)}\bigg|\mathcal{F}_t\right] = \mathbb{E}^{\mathcal{M}(B(\cdot))}\left[\frac{\mathcal{X}}{B(T)}\bigg|\mathcal{F}_t\right] \tag{A.6}$$

where $\mathbb{Q}$ is the (not necessarily unique) martingale measure for the market $B, S_1, \dots S_n$ with B as the numéraire.

A.5 Change of numéraire

The next lemma describes which change of measure turns the product of a martingale and a positive stochastic process into a new martingale.

Lemma A.6 (Change of numéraire) *Assume that $\mathcal{M}(S_1(\cdot))$ is a martingale measure for the numéraire S_1 (on $\mathcal{F}_T$) and assume that S_2 is a positive asset price process such that $S_2(t)/S_1(t)$ is a $\mathcal{M}(S_1(\cdot))$ martingale. Define $\mathcal{M}(S_2(\cdot))$ on $\mathcal{F}_T$ by the likelihood process*

$$L_1^2(t) = \frac{S_1(0)}{S_2(0)}\frac{S_2(t)}{S_1(t)}, \quad 0 \le t \le T. \tag{A.7}$$

Then $\mathcal{M}(S_2(\cdot))$ is a martingale measure for the numéraire S_2.

Proofs of theorems A.4, A.5 and lemma A.6 can be found in Björk [24].

Remark A.7 *Assuming S-dynamics of the form*

$$dS_i(t) = \alpha_i(t)S_i(t)dt + S_i(t)\sigma_i(t)dW^{\mathbb{P}}, \quad i = 1, 2\,,$$

Itō's formula applied to (A.7) gives the Girsanov kernel for the transition from $\mathcal{M}(S_1(\cdot))$ to $\mathcal{M}(S_2(\cdot))$:

$$\phi_1^2(t) = \sigma_2(t) - \sigma_1(t).$$

A **zero-coupon bond** is an asset that pays one unit currency at maturity T.

Definition A.8 *The risk-neutral martingale measure that arises from choosing the zero-coupon bond with maturity T as the numéraire in lemma A.6 is called the* ***T-forward measure*** $\mathcal{M}(p(\cdot,T))$.

The change of numéraire lemma A.6 provides us with a Radon-Nikomdym derivative

$$L_{\mathcal{M}(B(\cdot))}^{\mathcal{M}(p(\cdot,T))} = \frac{B(0)}{p(0,T)}\frac{p(s,T)}{B(s)} \tag{A.8}$$

relating $\mathcal{M}(B(\cdot))$ to $\mathcal{M}(p(\cdot,T))$. It follows that

$$\begin{aligned}
\frac{\Pi(s)}{p(s,T)} &= \frac{B(s)}{p(s,T)}\frac{\Pi(s)}{B(s)} \overset{\text{thm. A.5}}{=} \frac{B(s)}{p(s,T)}\mathbb{E}^{\mathcal{M}(B(\cdot))}\left[\frac{\Pi(t)}{B(t)}\middle|\mathcal{F}_s\right] \\
&= \frac{\frac{B(0)}{p(0,T)}\mathbb{E}^{\mathcal{M}(B(\cdot))}\left[\frac{\Pi(t)}{B(t)}\middle|\mathcal{F}_s\right]}{\frac{p(s,T)}{p(0,T)}\frac{B(0)}{B(s)}} \overset{\text{(A.8)}}{=} \frac{\mathbb{E}^{\mathcal{M}(B(\cdot))}\left[L_{\mathcal{M}(B(\cdot))}^{\mathcal{M}(p(\cdot,T))}\frac{\Pi(t)}{p(t,T)}\middle|\mathcal{F}_s\right]}{L_{\mathcal{M}(B(\cdot))}^{\mathcal{M}(p(\cdot,T))}} \\
&\overset{\text{Bayes' form.}}{=} \mathbb{E}^{\mathcal{M}(p(\cdot,T))}\left[\frac{\Pi(t)}{p(t,T)}\middle|\mathcal{F}_s\right].
\end{aligned}$$

In particular, as $p(T,T) = 1$:

Lemma A.9 *For any T-claim $\mathcal{X}$*

$$\frac{\Pi(t;\mathcal{X})}{p(t,T)} = \mathbb{E}^{\mathcal{M}(p(\cdot,T))}\left[\frac{\Pi(T;\mathcal{X})}{p(T,T)}\middle|\mathcal{F}_t\right] = \mathbb{E}^{\mathcal{M}(p(\cdot,T))}\left[\mathcal{X}|\mathcal{F}_t\right].$$

Lemma A.10 $\mathcal{M}(B(\cdot))$ *is equal to* $\mathcal{M}(p(\cdot,T))$ *iff. the interest rate r is deterministic.*

Proof The two measures $\mathcal{M}(p(\cdot,T))$ and $\mathcal{M}(B(\cdot))$ being equal implies their Radon-Nikodym derivative to be one. From equation (A.8) it can be seen that this is equivalent with the relation $p(t,T) = p(0,T)B(t)$ to hold for all $0 \le t \le T$. In an arbitrage free market with stochastic interest rate such a relation cannot hold since it implies you can always exchange a position in the bond for a position in the money account and vice versa, at no cost. If, on the other hand, the interest rate is deterministic, then $p(t,T) \neq p(0,T)B(t)$ for some t clearly leads to arbitrage opportunities.

Example A.11 (Call option with stochastic interest rate) *Given a financial market with stochastic short rate r and a strictly positive asset price process $S(t)$, consider a European call on S with maturity T and strike K, i.e. a T-claim $\mathcal{X} = \max\{0, S(T) - K\}$. Using the General pricing formula (Theorem A.5) and changing the numéraire we obtain the following expression for the option price:*

$$\mathbf{CALL}(S,T;0,K) = B(0)\,\mathbb{E}^{\mathcal{M}(B(\cdot))}\left[\frac{\max\{\,0,\ S(T)-K\,\}}{B(T)}\right]$$

$$
\begin{aligned}
&= B(0)\,\mathbb{E}^{\mathcal{M}(B(\cdot))}\left[\frac{S(T)}{B(T)}\mathbb{1}(S(T)>K)\right] - K\,B(0)\,\mathbb{E}^{\mathcal{M}(B(\cdot))}\left[\frac{1}{B(T)}\mathbb{1}(S(T)>K)\right] \\
&= S(0)\,\mathbb{E}^{\mathcal{M}(S(\cdot))}\left[\frac{S(T)}{S(T)}\mathbb{1}(S(T)>K)\right] - K\,p(0,T)\,\mathbb{E}^{\mathcal{M}(p(\cdot,T))}\left[\frac{1}{p(T,T)}\mathbb{1}(S(T)>K)\right] \\
&= S(0)\,\mathbb{E}^{\mathcal{M}(S(\cdot))}\left[\mathbb{1}(S(T)>K)\right] - K\,p(0,T)\,\mathbb{E}^{\mathcal{M}(p(\cdot,T))}\left[\mathbb{1}(S(T)>K)\right], \qquad \text{(A.9)}
\end{aligned}
$$

where $\mathbf{CALL}(S,T;t,K)$ *denotes the price at time* t *of a call option on* S *with maturity* T *and strike* K*.*

Under the assumptions of the Black-Scholes model (asset and money account are driven by the same Wiener process, deterministic interest rate) equation (A.9) reduces to the Black-Scholes formula (A.4).

If interest rate is stochastic, but $S(t)/p(t,T)$ has deterministic volatility, equation (A.9) can still be evaluated analytically:

Lemma A.12 (Geman–El Karoui–Rochet) *Under the assumption that the process* $\frac{S(t)}{p(t,T)}$ *has* ***deterministic volatility*** $\sigma_{S,T}(t)$*, equation (A.9) reduces to*

$$
\mathbf{CALL}(S,T;0,K) = S(0)N[d_1] - Kp(0,T)N[d_2], \qquad \text{(A.10)}
$$

where

$$
\begin{aligned}
d_1 &= d_2 + \widehat{\sigma}_{S,T}(T)\sqrt{T}, \qquad \text{(A.11)} \\
d_2 &= \frac{\log\left(\frac{S(0)}{Kp(0,T)}\right) - \frac{1}{2}\widehat{\sigma}^2_{S,T}(T)\,T}{\widehat{\sigma}_{S,T}(T)\,\sqrt{T}}, \qquad \text{(A.12)} \\
\widehat{\sigma}^2_{S,T}(T) &= \frac{1}{T}\int_0^T ||\sigma_{S,T}(t)||^2 dt. \qquad \text{(A.13)}
\end{aligned}
$$

B The assumption of lognormality

Consider the Black-Scholes model of Example A.3. The solution of SDE (A.2) is given by

$$S(t) = S(0)\exp\{(\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)\}.$$

Thus, under Black-Scholes, the price $S(t)$ is lognormally distributed for all t. In practise though, the distribution appears to be different. In particular, the tails of the distribution which represent the probability of extreme values of the stock are heavier than assumed in Black-Scholes.

Two possible solutions to this problem are described in the following sections. A common approach is to parametrize the **implied distribution**, i.e. the (possibly non-normal) distribution that prices back the options observed in the market. In the particular case of the Black-Scholes model, there exists a one-to-one relationship between the implied distribution and the **equivalent normal volatility** or **implied volatility** that (given the returns being normal) leads to the correct market prices for calls and puts. This approach may seem dubious as in fact one 'cancels out' the effect of a wrong model by adapting the volatility. Its popularity is mainly due to normal distributions being analytically easy to handle.

Another solution is to use a different distribution for modelling the asset returns. The class of stable distributions which is described in Section B.2 allows for fat tails and contains normal distributions as a subset.

B.1 Implied distribution

Breeden and Litzenberger [26] showed that specifying all call option prices (i.e. for every possible strike) completely determines the terminal distribution of the asset underlying the options:

$$-\,e^{r(T-t)}\frac{\partial\mathbf{CALL}(S,t;T,K)}{\partial K} = \mathcal{M}(B(\cdot))\;(S(T)\geq K\,|\,\mathcal{F}_t). \tag{B.1}$$

Similarly, for the price of a put option:

$$e^{r(T-t)}\frac{\partial\mathbf{PUT}(S,t;T,K)}{\partial K} = \mathcal{M}(B(\cdot))\;(S(T)\leq K\,|\,\mathcal{F}_t). \tag{B.2}$$

Given prices of options observed in the market, these relations specify the market implied distribution. Dupire [27] showed that under risk neutrality, there is a unique **local volatility function** $\sigma(t,T)$ consistent with the implied distribution.

Analogously, the **implied volatility** is the equivalent normal volatility that prices back market observed quotes using the Black-Scholes formula (A.4). It can be shown that specifying these implied volatilities for every strike completely determines the distribution:

$$
\begin{aligned}
\mathbb{P}[S(T) < K] &\overset{(B.1)}{=} 1 + p(0,T)^{-1}\frac{\partial}{\partial K}\mathbf{CALL}(T,K;t=0) \\
&\overset{(A.10)}{=} 1 + \frac{\partial}{\partial K}\Big(F^T(0)\Phi(d_1) - K\Phi(d_2)\Big) \\
&= 1 - \Phi(d_2) - K\frac{\partial}{\partial K}\Phi(d_2) + \frac{\partial}{\partial K}\Phi(d_1)F^T(0) \\
&= \Phi(-d_2) - K\,\phi(d_2)\,\frac{\partial d_2}{\partial K} + F^T(0)\,\phi(d_1)\,\Big(\frac{\partial d_2}{\partial K} + \sqrt{T}\frac{\partial\widehat{\sigma}}{\partial K}\Big) \\
&= \Phi(-d_2) + F^T(0)\,\sqrt{T}\,\phi(d_1)\,\frac{\partial\widehat{\sigma}}{\partial K} + \Big(F^T(0)\phi(d_1) - K\phi(d_2)\Big)\,\frac{\partial d_2}{\partial K}
\end{aligned}
$$

where d_1, d_2 are given by equations (A.11) and (A.12). The last term can be seen to be zero:

$$
\begin{aligned}
\phi(d_2) &= \exp(-\frac{1}{2}d_1^2) \;=\; \exp\left(-\frac{1}{2}(d_2^2 + 2\,\widehat{\sigma}\,\sqrt{T}\,d_2 + \widehat{\sigma}^2\,T)\right) \\
&= \exp(-\frac{1}{2}d_2^2)\,\exp(-\widehat{\sigma}\,\sqrt{T}\,d_2 - \frac{1}{2}\widehat{\sigma}^2\,T) \;=\; \frac{F^T(0)}{K}\phi(d_1)\,.
\end{aligned}
$$

We are thus left with

$$
\mathbb{P}[S(T) < K] = \Phi(-d_2) + F^T(0)\,\sqrt{T}\,\phi(d_1)\,\frac{\partial\widehat{\sigma}}{\partial K}\,. \tag{B.3}
$$

B.2 Stable distributions

Stable distributions are implicitly defined by the property that sums of independent and identically distributed copies of a stable random variable follow the same distribution up to multiplication by a positive constant and a constant shift. Normal distributions for example satisfy this property.

An equivalent definition [28] is to call a random variable X stable if and only if it is equal in distribution to $aZ + b$ where $a > 0$, $b \in \mathcal{R}$, $0 < \alpha \le 2$, $-1 \le \beta \le 1$ and Z is a random variable with characteristic function

$$
\mathbb{E}\left[e^{iuZ}\right] = \begin{cases} \exp(-|u|^\alpha[1 - i\beta\tan(\frac{\pi\alpha}{2}\text{ sign } u)\,]\,) & \alpha \ne 1, \\ \exp(-|u|\;[1 + i\beta\frac{2}{\pi}(\text{ sign } u\,)\log|u|\,]\,) & \alpha = 1. \end{cases}
$$

The limiting behaviour of the tails is determined by the index of stability α. The special case $\alpha = 2$ corresponds to the normal distribution where the tails behave as an exponential function. If $\alpha < 2$ the tails are heavier than exponential, i.e. $x^\alpha\ \mathbb{P}[X > x] \to c^+$ and $x^\alpha\ \mathbb{P}[X < -x] \to c^-$ as $x \to \infty$ where $c^+, c^- \in \mathcal{R}$ [28, Theorem 1.12].

Stable distributions are the only distributions that can occur as limiting cases of sums of i.i.d. random variables [28, Theorem 1.20]. This result is known as the Generalized Central Limit Theorem. The generalization lies in the fact that the random variables to which this theorem applies need not to have a finite second moment as is the case for the standard Central Limit Theorem.

The property of heavy tails led to the use of stable distributions in financial returns modelling, initially by Mandelbrot in his 1963 article on cotton prices [29, Chapter E14]. A drawback of this approach is the fact that the second moment of stable random variables for $\alpha < 2$ is infinite and many consider this to be an unrealistic assumption for asset returns. Also, the index of stability which determines the tail behaviour is invariant under (i.i.d.) addition leading to overestimation of the tails on larger timescales [30, Figure 2]. In the literature, solutions have been proposed to repair both the convergence issue (slower than normal, faster than stable) and the finite variance — for instance by means of truncated Lévy flights [31]. Analytically though, these appoaches are much less attractive than the normal distribution.

References

[1] Y. Malevergne and D. Sornette. Testing the gaussian copula hypothesis for financial assets dependences. Working paper, Institute of Geophysics and Planetary Physics, 2001.

[2] Roger B. Nelsen. *An Introduction to Copulas*. Springer, 1999.

[3] B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*. Elsevier Science, 1983.

[4] M. Scarsini. On measures of concordance. *Stochastica*, 8:201–218, 1984.

[5] U. Cherubini, E. Luciano, and W. Vecchiato. *Copula Methods in Finance*. Wiley, 2004.

[6] H. Joe. Multivariate concordance. *Journal of Multivariate Analysis*, 35:12–30, 1990.

[7] R. Schmidt. Tail dependence for elliptically contoured distributions. *Math. Methods of Operations Research*, 55:301–327, 2002.

[8] A.W. Marshall and I. Olkin. Families of multivariate distributions. *Journal of the American Statistical Association*, 83:834–841, 1988.

[9] C. H. Kimberling. A probabilistic interpretation of complete monotonicity. *Aequationes Math.*, 10:152–164, 1974.

[10] W. Feller. *An introduction to probability theory and its applications, vol. 2*. Wiley, 1971.

[11] H. Joe. *Multivariate models and dependence concepts*. Chapman and Hall, 1997.

[12] Jäckel P. A practical method for the valuation of a variety of hybrid products. ICBI conference slides, 2005.

[13] Hagan P.S., Kumar D., Lesniewski A.S., and Woodward D.E. Managing smile risk. *Wilmott Magazine*, 3:84–108, 2002.

[14] Chambers J.M., Mallows C.L., and Stuck B.W. A method for simulating stable random variables. *J. of the Am. Stat. Assoc.*, 71, No. 354:340–344, 1984.

[15] J. Shao. *Mathematical Statistics*, pages 248–254. Springer-Verlag, 1999.

[16] H. Joe and J.J. Xu. The estimation method of inference functions for margins for multivariate models. Technical Report 166, Dept. of Statistics University of British Columbia, 1996.

[17] F. Dellaert. The expectation maximization algorithm. GVU Center, College of Computing, Georgia Tech, 2002.

[18] J. Bilmes. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. Tutorial, International Computer Science Institute, Berkeley, 1997.

[19] R.A. Redner and Walker H.F. Mixture densities, maximum likelihood and the em algorithm. *SIAM Review*, 26, No. 2, 1984.

[20] Fermanian J.D. and Scaillet O. Some statistical pitfalls in copula modeling for financial applications. Research Paper, International Center for Financial Asset Management and Engineering, 2004.

[21] Rapuch G. and Roncalli T. Some remarks on two-asset options pricing and stochastic dependence of asset prices. Working paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais, 2001.

[22] L. Hu. Dependence patterns across financial markets: a mixed copula approach. *Applied Financial Economics*, 16:717–729, 2004.

[23] J.M. Steele. *Stochastic Calculus and Financial Applications*. Springer, 2000.

[24] T. Björk. *Arbitrage Theory in Continuous Time*. Oxford University Press, 2004.

[25] M. Baxter and A. Rennie. *Financial Calculus, An introduction to derivatives pricing*. Cambridge, 1996.

[26] D. Breeden and R. Litzenberger. Prices of state-contingent claims explicit in option prices. *Journal of Business*, 51:621–651, 1978.

[27] B. Dupire. Pricing with a smile. *Risk magazine*, 7:18–20, 1994.

[28] J. P. Nolan. *Stable Distributions - Models for Heavy Tailed Data*. Birkhäuser, Boston, 2007. In progress, Chapter 1 online at academic2.american.edu/∼jpnolan.

[29] B.B. Mandelbrot. *Fractals and scaling in finance*. Springer, 1997.

[30] Cont R., Potters M., and Bouchaud J.P. Scaling in stock market data: stable laws and beyond. Proceedings of the Les Houches workshop, France, 1997.

[31] Andrew Matacz. Financial modeling and option theory with the truncated levy process. Technical Report 97-28, University of Sydney, 1997.