Modeling tail dependence using copulas — literature review

Jan de Kort

March 15, 2007

Contents

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"). If the corn price is high, this is possibly due to unfavourable weather conditions, also causing the wheat harvest to be bad. High corn prices thus are likely to be observed together with high wheat prices. This phenomenon is called (upper) tail dependence.

In commodities 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 4.2.1 it will be shown that the Gaussian copula does not have tail dependence. This may cast some doubt on the appropriateness of this model in case of the corn and wheat option, for the probability of both crops having high prices will be underestimated by the Gaussian copula. This leads to underpricing of the best-of contract.

Malevergne and Sornette [1] show that a Gaussian copula might 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.

Still, copulas provide a convenient way to model the dependence between assets, since the marginal distributions of the underlyings can be dealt with separately so that their properties (such as volatility smile) can be preserved.

The central question in this project is **how to incorporate tail dependence in the pricing of hybrid products**. A first step would be to consider best-of contracts on two underlyings.

This literature review seeks to give an overview of the theory involved in answering the above question. Sections 1 and 2 explain what copulas are and how they relate to multivariate distribution functions. In section 3 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 4 summarizes the properties of a number of well-known parametric families of copulas. Some calibration techniques are outlined in section 5. Section 6 finally describes how prices of hybrid contracts (e.g. digital options, best-of contracts) can be expressed in terms of copulas.

Appendix A provides a brief introduction to derivatives pricing, in particular the role that martingales play in this. Since the marginal distributions of the hybrid derivatives should be chosen as to incorporate volatility smile, a short note on implied volatility / implied distribution can be found in appendix B.

1 Bivariate Copulas

This section introduces copulas and describes how they relate to multivariate distributions (i.e. Sklar's theorem, section 1.1). Section 1.2 introduces the Fréchet-Hoeffding upper and lower bounds for copulas. It is also explained what it means for a multivariate distribution if its copula is maximal or minimal. Finally, in section 1.3, survival copulas are defined which will be useful in the discussion of tail dependence later on.

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

Definition 1.1 *Let $\varnothing \neq S_1, S_2 \subset \overline{\mathcal{R}}$ and let H be a $S_1 \times S_2 \to \mathcal{R}$ function. The **H-volume** of $B = [x_1, x_2] \times [y_1, y_2]$ 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 1.2 *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 1.3 *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 $-\infty$.

As H is 2-increasing we have, from definition 1.1,

$$H(x_2, y_2) - H(x_1, y_2) \geq H(x_2, y_1) - H(x_1, y_1) \tag{1}$$

and

$$H(x_2, y_2) - H(x_2, y_1) \geq H(x_1, y_2) - H(x_1, y_1) \tag{2}$$

for every $[x_1, x_2] \times [y_1, y_2] \subset S_1 \times S_2$. By setting $x_1 = a_1$ in (1) and $y_1 = a_2$ in (2) it can be seen that

Lemma 1.4 *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*

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

From lemma 1.4 it follows that (1) and (2) also hold in absolute value. Adding up these inequalities and applying the triangle inequality (Figure 1) yields in particular

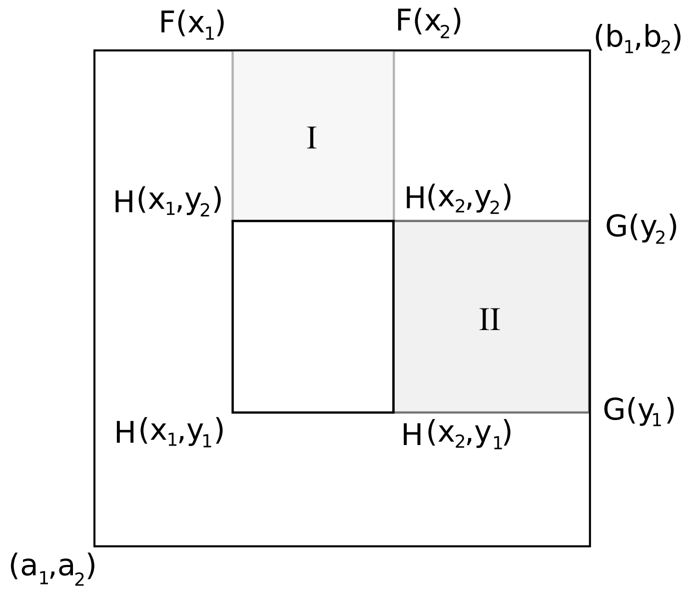

Figure 1: Schematic proof of lemma 1.5. Apply 2-increasingness to rectangle I and II and combine the resulting inequalities via the triangle inequality. For the absolute value bars, use that H is nondecreasing in both arguments (lemma 1.4).

Lemma 1.5 *For any grounded, 2-decreasing 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 1.6 *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 1.7 *A (two dimensional)* ***copula*** *is a subcopula whose domain is* $[0, 1]^2$.

Remark 1.8 *Note that reformulating lemma 1.5 in terms of subcopulas immediately leads to the Lipschitz condition*

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

guarantying continuity of (sub)copulas.

Definition 1.9 *The* ***density*** *associated with a copula* C *is*

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

Definition 1.10 *The* ***absolutely continuous component*** A_C *and the and the* ***singular component*** S_C *of the density are defined as*

$$\begin{aligned} A_C(u,v) &= \int_0^u \int_0^v \frac{\partial^2 C(s,t)}{\partial u \partial v}\, ds\, dt, \\ S_C(u,v) &= C(u,v) - A_C(u,v). \end{aligned}$$

1.1 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 seperate 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 1.11 *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*

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

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

Definition 1.12 *Let* X *be a random variable. The* ***cumulative distribution function*** *(CDF) of* X *is*

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

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

Definition 1.13 *If the derivative of the CDF of* X *exists, it is called the* ***probability density function*** *(pdf) of* X*.*

Definition 1.14 *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 1.15 *A random variable is said to be* ***continuous*** *if its CDF is continuous.*

Lemma 1.16 *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

$$H(x,y) = C'(F(x), G(y)) \tag{3}$$

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

Proof 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 (3). Suppose to the contrary that $C_1'(u,v) \neq C_2'(u,v)$ are both consistent with (3), i.e. there exist $(x_1,y_1),(x_2,y_2) \in \overline{\mathcal{R}}^2$ such that

$$C_1'(u,v) = C_1'(F(x_1), G(y_1)) = H(x_1,y_1),$$

$$C_2'(u,v) = C_2'(F(x_2), G(y_2)) = H(x_2,y_2).$$

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 1.5 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).

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 1.17 *The converse of lemma 1.16 also holds: every H defined by (3) is a joint distribution. This follows from the properties of a subcopula.*

Theorem 1.18 (Sklar's theorem) *Let H be a joint distribution function with margins F and G. Then there exists a unique 2-copula C such that for all $(x, y) \in \overline{\mathcal{R}}^2$*

$$H(x, y) = C(F(x), G(y)). \quad (4)$$

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 (4) is a joint distribution function with margins F and G.

Proof Lemma 1.16 provides us with a unique subcopula C' satisfying (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 1.17 for copulas. $\square$

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

Theorem 1.19 (C invariant under increasing transformation 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 $\alpha(X) \sim F_\alpha$ and $\beta(Y) \sim G_\beta$ have copula $C_{\alpha\beta} = C$.*

Proof

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

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

Theorem 1.20 *The continuous random variables X and Y are independent iff. their copula is $C^\perp(u, v) = uv$.*

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

1.2 Fréchet-Hoeffding bounds

In this section we will show the existence of a maximal and a minimal bivariate copula, usually refered 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.

Theorem 1.21 *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.*

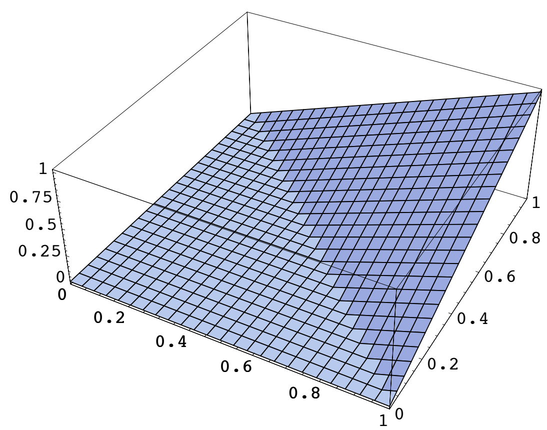

Figure 2: Fréchet-Hoeffding lower bound

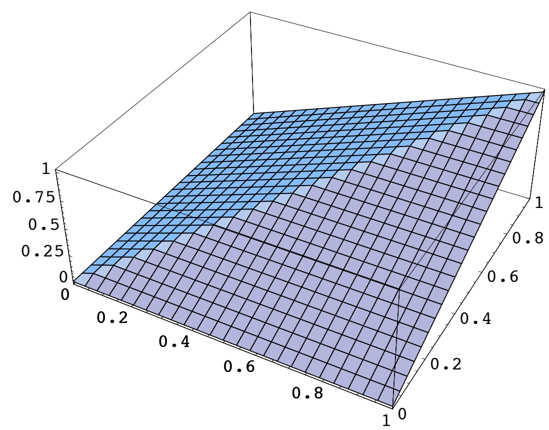

Figure 3: Fréchet-Hoeffding upper bound

Proof From lemma 1.4 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 and 3. The remaining part of this section is devoted to the question under what condition these bounds are attained.

Definition 1.22 *A set $S := S_1 \times S_2 \subset \overline{\mathcal{R}}^2$ 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 4.

Definition 1.23 *The* ***support*** *of a distribution function H is the complement of the union of all open subsets of $\mathcal{R}^2$ with H-measure zero.*

Remark 1.24 *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 1.23.*

Let X and Y be random variables with joint distribution H and continuous margins $F : S_1 \to \mathcal{R}$ and $G : S_1 \to \mathcal{R}$. 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

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

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 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].

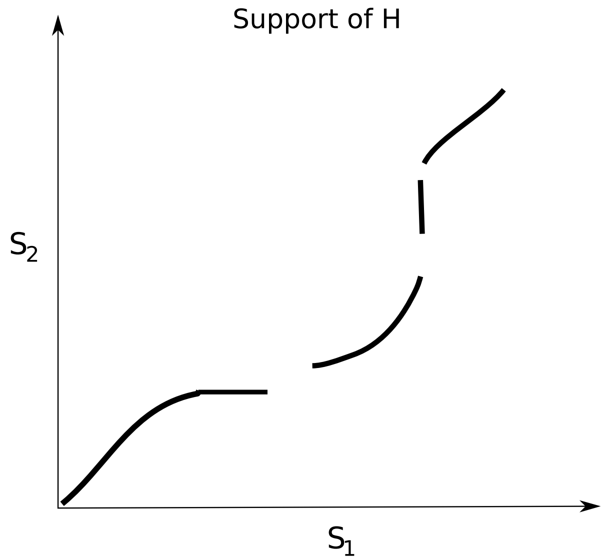

Figure 4: Example of a nondecreasing set.

Theorem 1.25 *Let X and Y be random variables with joint distribution function H.*

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

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

Remark 1.26 *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 1.25 implies the support of H to be an almost surely increasing (decreasing) set iff. H is equal to the upper (lower) Fréchet-Hoeffding bound.

Remark 1.27 *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'.*

1.3 Survival copula

Every bivariate copula has a survival copula associated with it that gives the probability of two random variables both to exceed a certain value.

Definition 1.28 *The **survival copula** associated with the copula C is*

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

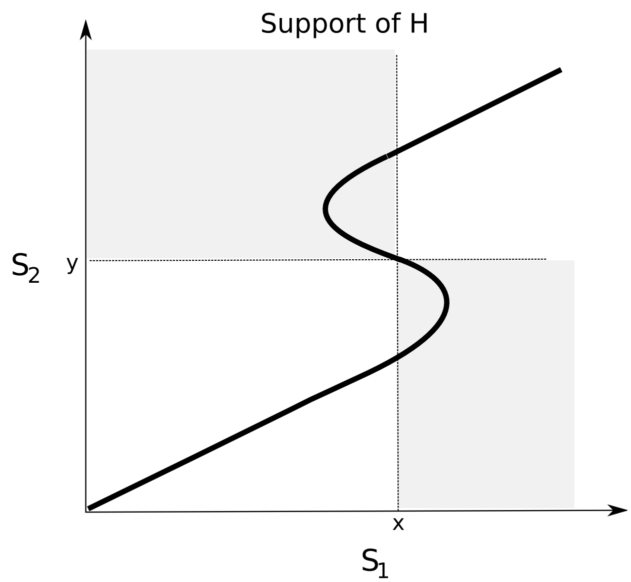

Figure 5: In case of non-perfect positive dependence, the shaded area always contains points with nonzero probability.

Indeed, $\overline{C}$ is a copula:

$$\begin{aligned}
\overline{C}(0,v) &= 0+v-1+C(1,1-v) = v-1+1-v = 0,\\
\overline{C}(1,v) &= 1+v-1+C(0,v) = v+0 = v,
\end{aligned}$$

The other verifications are similar.

Consider two random variables $X \sim F$, $Y \sim G$ with copula C and joint distribution function H, then

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

2 Multivariate Copulas

The notion of copulas, introduced in section 1, 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 2.1 *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 2.2 *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 2.3 *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 $\text{ver([x,y])} \in S_1 \times \ldots \times S_n$ is nonnegative:*

$$\sum_{w \in \text{ver}([x,y])} \text{sgn(w)} H(w) \geq 0 \tag{5}$$

Definition 2.4 *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 2.5 *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 2.6 *An n-dimensional subcopula for which $S_1 \times \ldots \times S_n = I^n$ is an **n-dimensional copula**.*

2.1 Sklar's theorem

Theorem 2.7 (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 $u \in \overline{\mathcal{R}}^n$*

$$H(u_1, \ldots, u_n) = C(F(u_1), \ldots, F(u_n)). \tag{6}$$

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 (6). is a joint distribution function with margins $F_1, \ldots, F_n$.

2.2 Fréchet-Hoeffding bounds

Theorem 2.8 *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 (5). 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 Dependence

The dependence structure between random variables is completely described by their joint distribution function. 'Benchmarks' like linear correlation 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.

One could think of measures of association as 'one dimensional projections of the dependence structure onto the real line'. Consider for instance Scarsini's [4] definition:

"*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.*"

From this definition it is clear that there exists some freedom in how to define the 'extent to which X and Y cluster around the graph of a function'. In the following, some of the standard interpretations of this freedom will be described.

Section 3.1 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 3.2 and 3.3 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 3.5(6) — and copulas are invariant under monotone transformations of the random variables.

Finally, in section 3.4 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.

3.1 Linear correlation

Definition 3.1 *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 3.2 *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

$$\mathbb{E}\left[X^n\right] = \int_0^1 x^n\,dx = \frac{1}{1+n}. \tag{7}$$

The linear correlation between X and Y is

$$\begin{aligned}
\rho &= \frac{\mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]}{(\mathbb{E}[X^2] - \mathbb{E}[X]^2)^{\frac{1}{2}}(\mathbb{E}[Y^2] - \mathbb{E}[Y]^2)^{\frac{1}{2}}} \\
&= \frac{\mathbb{E}\left[X^{n+1}\right] - \mathbb{E}[X]\,\mathbb{E}[X^n]}{(\mathbb{E}[X^2] - \mathbb{E}[X]^2)^{\frac{1}{2}}(\mathbb{E}[X^{2n}] - \mathbb{E}[X^n]^2)^{\frac{1}{2}}} \\
&\overset{(7)}{=} \frac{\sqrt{3+6n}}{2+n}.
\end{aligned}$$

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

Corollary 3.3 *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.

3.2 Measures of concordance

Definition 3.4

*(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*

$$C_1(u) \le C_2(u) \quad \text{and} \quad \overline{C}_1(u) \le \overline{C}_2(u) \quad \text{for all } u \in I^m. \tag{8}$$

In the remaining part of this section we will only consider bivariate copulas. Part (ii) of definition 3.4 is then equivalent to $C_1(u, v) \le C_2(u, v)$ for all $u \in I^2$.

Definition 3.5 *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. *$-1 \le \kappa_{X,Y} \le 1$, $\kappa_{X,X} = 1$, $\kappa_{-X,X} = -1$,*
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 3.4 and 3.5?

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 1.19 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 3.4 and 3.5 are related.

Summarizing,

Lemma 3.6

(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 exactly opposite to the conclusions in corollary 3.3 on the linear correlation coefficient. The linear correlation coefficient 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.

3.2.1 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):

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

In case (X_1, Y_1) and (X_2, Y_2) are **iid.** 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

$$Q = 4 \iint_{I^2} C_2(u, v)\, dC_1(u, v) - 1. \tag{9}$$

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, (9) gives Kendall's tau for the iid. random vectors (X_1, Y_1), (X_2, Y_2) with copula C.

Furthermore it can be shown that

$$\tau = 1 - 4 \iint_{I^2} \frac{\partial C(u, v)}{\partial u} \frac{\partial C(u, v)}{\partial v}\, du\, dv. \tag{10}$$

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 (10), consider

Lemma 3.7 $\tau_C = \tau_{\overline{C}}$.

Proof

$$\begin{aligned} \tau_{\overline{C}} &= 1 - 4 \iint_{I^2} \frac{\partial \overline{C}}{\partial u} \frac{\partial \overline{C}}{\partial v}\, du\, dv \\ &= 1 - 4 \iint_{I^2} [1 - \frac{\partial C}{\partial u}][1 - \frac{\partial C}{\partial v}]\, du\, dv \\ &= \tau_C - 4 \iint_{I^2} [1 - \frac{\partial C}{\partial u} - \frac{\partial C}{\partial v}]\, du\, dv. \end{aligned} \tag{11}$$

The second term of the integrand of (11) 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 (11) yields the lemma. □

Scarsini [4] proves axioms (1)–(7) of definition 3.5 hold for Kendall's tau.

The next lemma does not hold for general concordance measures.

Lemma 3.8 *Let H be a joint distribution with copula C.*

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

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 1.25. Step (iii) follows from substitution of C^+ in formula (9) and straightforward calculation. This step in fact also follows from axiom (6) in definition 3.5.

It remains to show that $\tau = 1$ implies H having nondecreasing support. Suppose therefore that

$$1 = \tau = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Clearly,

$$\mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0] = 0. \tag{12}$$

Now suppose that (x_1, y_1) and (x_2, y_2) are disconcordant and lie in the support of H. Integrating (12) over I^2 yields

$$\begin{aligned} 0 &= \iint_{I^2} \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0] \, du \, dv \\ &= \iint_{I^2} \mathbb{P}[(X_2 - x_1)(Y_2 - y_1) < 0 | (X_1, Y_1) = (x_1, y_1)] \, dH(u, v) \\ &= \iint_{I^2} \Big\{ \mathbb{P}[X_2 > x_1, Y_2 < y_1 | (X_1, Y_1) = (x_1, y_1)] + \mathbb{P}[X_2 < x_1, Y_2 > y_1 | (X_1, Y_1) = (x_1, y_1)] \Big\} \, dH(u, v). \end{aligned}$$

It follows that there is no probability mass in the regions $\{(u, v) : u > x_1, v < y_1\}$ and $\{(u, v) : u < x_1, v > y_1\}$. In particular

$$\mathbb{P}[\mathcal{B}_r(x_1, y_1)] = 0,$$

where $r := \frac{1}{2} | \min\{x_1 - x_2, y_1 - y_2\}|$ and $\mathcal{B}_r(x, y)$ denotes an open 2-ball with radius r and centre (x, y). Apparently (x_1, y_1) is not in the complement of the union of open sets having zero probability and therefore not in the support of H. This contradicts our assumption and proves (i). □

3.2.2 Spearman's rho

Let (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3) be iid. 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 $(\underbrace{X_2, Y_3}_{\text{Independent}})$:

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

Note that X_2 and Y_3, being independent, have copula $C^{\perp}$. By (9), three times the concordance difference between C and $C^{\perp}$ is

$$\begin{aligned} \rho_S &= 3\left(4 \iint_{I^2} C(u,v) dC^{\perp}(u,v) - 1 \right) \\ &= 12 \iint_{I^2} C(u,v)\, du\, dv - 3. \end{aligned} \tag{13}$$

Spearman's rho satisfies the axioms in definition 3.5 (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 $\text{Var}[F(X)] = \text{Var}[G(Y)] = 1/12$. We thus have

$$\begin{aligned} \rho_S &\overset{(13)}{=} 12\, \mathbb{E}[F(X), G(Y)] - 3 \\ &= \frac{\mathbb{E}[F(X), G(Y)] - (1/2)^2}{1/12} \\ &= \frac{\mathbb{E}[F(X), G(Y)] - \mathbb{E}[F(X)]\, \mathbb{E}[G(Y)]}{(\text{Var}[F(X)]\text{Var}[G(Y)])^{\frac{1}{2}}} \\ &= \frac{\text{Cov}[F(X), G(Y)]}{(\text{Var}[F(X)]\text{Var}[G(Y)])^{\frac{1}{2}}}. \end{aligned}$$

Cherubini et al. [5] states that for Spearman's rho a statement similar to lemma 3.8 holds: $C = C^{\pm}$ iff. $\rho_S = \pm 1$.

3.2.3 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^-. Using (9) this reads

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

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

3.3 Measures of dependence

Definition 3.9 *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. X and Y are strictly monotone functions of eachother,*
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, or, in the words of Scarsini [4]:
"*[..] dependence is a matter of assocation with respect to a (strictly) monotone function (indifferently increasing or decreasing). [..] Concordance, on the other hand, takes into account the kind of monotonicity [..] the maximum is attained when a strictly increasing relation exists [..] the minimum [..] when a relation exists that is strictly monotone decreasing.*"

3.3.1 Schweizer and Wolff's sigma

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

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

Nelsen [2] shows this association measure to satisfy the properties of definition 3.9.

3.4 Tail dependence

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

Definition 3.10 *Given two random variables $X \sim F$ and $Y \sim G$ with copula C, define the coefficients of tail dependency*

$$\lambda_L := \lim_{u \downarrow 0} \mathbb{P}[F(X) < u | G(Y) < u] = \lim_{u \downarrow 0} \frac{C(u,u)}{u}, \tag{14}$$

$$\lambda_U := \lim_{u \uparrow 1} \mathbb{P}[F(X) > u | G(Y) > u] = \lim_{u \uparrow 1} \frac{1 - 2u + C(u,u)}{1-u}. \tag{15}$$

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

The interpretation of the coefficients of tail dependency is that it measures the probability of two random variables both taking extreme values.

Lemma 3.11 *Denote the lower (upper) coefficient of tail dependency of the survival copula $\overline{C}$ by $\overline{\lambda}_L$ $(\overline{\lambda}_U)$, then*

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

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 3.12 *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 $(0,0)$ and $(1,1)$, this is a sufficient condition for the limits (14) and (15) to exist:

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

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

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

3.5 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 iid. 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 3.4(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.

4 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 4.1) 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 4.2 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 4.3) 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 4.4 an extension of the class of Archimedean copulas will be discussed that allows for several distinct bivariate margins.

4.1 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.

4.2 Elliptical distributions

Elliptical distributions are distributions whose density function (if it exists) equals

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

where Σ (dispersion) is a symmetric positive semi-definite matrix, $\mu \in \mathcal{R}^n$ (location) and g (density generator) is a $[0,\infty) \to [0,\infty)$ function.

Taking $g(x) = \frac{1}{2\pi}\exp\{-\frac{x}{2}\}$ yields the Gaussian distribution (Section 4.2.1) and $g(x) = \left(1+\frac{tx}{\nu}\right)^{-\frac{2+\nu}{2}}$ leads to a Student's t distribution with ν degrees of freedom (Section 4.2.2).

Schmidt [7] 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 f is called **regularly varying with index** α if for every $t > 0$

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

In words: regularly varying functions have tails that behave like power functions.

Whether or not the generator being regularly varying is a necessary condition for tail dependence is still an open problem, but Schmidt proves that to have tail dependency, at least one bivariate 'margin' must be **O-regularly varying**, that is it must satisfy

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

for every $t \ge 1$.

4.2.1 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 characterisation of tail dependent elliptical distributions, since the density generator for the bivariate Gaussian distribution ($\rho \neq 1$) is not O-regularly varying:

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

4.2.2 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 $\nu > 0$.

4.3 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

$$C^A(u,v) = \phi^{[-1]}(\phi(u)+\phi(v)) \tag{16}$$

is a copula and is called the **Archimedean copula with generator** ϕ. The density of C^A is given by

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

4.3.1 One-parameter families

The Gumbel copula from example 3.12 is Archimedean with generator $\phi(u) = (-\log(u))^\theta$, $\theta \in [1,\infty)$. Some other examples are listed in table 1.

Table 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*	$\left(\max\{0, u^{-\theta}+v^{-\theta}-1\}\right)^{-\frac{1}{\theta}}$	$\frac{1}{\theta}(t^{-\theta}-1)$	$[-1,\infty)\backslash\{\varnothing\}$	$\frac{\theta}{\theta+2}$	$2^{-\frac{1}{\theta}}$	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*	$-\frac{1}{\theta}\log\left(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$	$-\log\frac{e^{-\theta t}-1}{e^{-\theta}-1}$	$(-\infty,\infty)\backslash\{\varnothing\}$		0	0

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

$$C^A(u,u)=\phi^{[-1]}(2\phi(u))<\phi^{[-1]}(\phi(u))=u. \tag{17}$$

As $C^+(u,u)=u$, inequality (17) 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.

4.3.2 Two-parameter families

Nelsen [2] shows that if ϕ is a stict 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 1.

4.3.3 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

$$(-1)^k \frac{d^k \phi^{-1}(t)}{dt^k} \geq 0 \quad \text{for all } t \in [0, \infty),\ \ k = 1, \ldots, n \tag{18}$$

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 (18). 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 (18) 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.

4.4 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

$$C_\phi(C_\psi(u,v), w) \tag{19}$$

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 (19) 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.

5 Calibration of Copulas from 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 $\{Y_t,\ t = 1, \ldots, T\}$.

5.1 Maximum likelihood method

Let $X_i \sim F_i$, $1 \leq i \leq n$, 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 to $u_1, u_2, \ldots, u_n$ sequentially yields the **canonical representation**

$$h(u_1, \ldots, u_n) = c(F_1(u_1), \ldots, F_n(u_n)) \prod_{i=1}^{n} f_i(u_i), \tag{20}$$

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 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

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

This θ also maximimes the log-likelihood

$$\log l(\theta) = \sum_{t=1}^{T} \log c(F_1(x_{1t}), \ldots, F_n(x_{nt});\ \theta) + \sum_{t=1}^{T} \sum_{i=1}^{n} \log f_i(x_{it};\ \theta). \tag{21}$$

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

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

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 [12]) that if the sample size increases, the subsequent MLEs converge to a limit. This property is called consistency.

5.2 IFM method

The log-likelihood (21) consists of two positive parts. Joe and Xu [13] 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.

$$\theta_1 = \underset{\theta_2}{\operatorname{argmax}} \sum_{t=1}^{T} \sum_{i=1}^{n} \log f_i(x_{it};\ \theta_1)$$

$$\theta_2 = \underset{\theta_1}{\operatorname{argmax}} \sum_{t=1}^{T} \log c(F_1(x_{1t}), \ldots, F_n(x_{nt});\ \theta_1, \theta_2)$$

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.

5.3 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:

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

6 Option Pricing with Copulas

In this section, existing pricing techniques for (bivariate) digital and rainbow options with dependent underlyings will be restated in terms of copulas. Following Cherubini et al. [5], a martingale approach will be used. An outline of martingale pricing is given in appendix A.4.

We will need a reformulation of Sklar's theorem in terms of conditional copulas. Conditional copulas are defined along the same lines as in section 1, details can be found in Patton [14].

Theorem 6.1 (Sklar's theorem for conditional copulas) *Let $X|\mathcal{F}_t \sim F(\cdot|\mathcal{F}_t)$, $Y|\mathcal{F}_t \sim G(\cdot|\mathcal{F}_t)$ be random variables with joint conditional distribution $H(X,Y|\mathcal{F}_t)$. Then there exists a copula C such that*

$$H(x,y|\mathcal{F}_t) = C(F(x|\mathcal{F}_t), G(y|\mathcal{F}_t) \mid \mathcal{F}_t).$$

Conversely, given two conditional distributions $F(\cdot|\mathcal{F}_t)$ and $G(\cdot|\mathcal{F}_t)$ the composition $C(F(\cdot|\mathcal{F}_t), G(\cdot|\mathcal{F}_t))$ is a conditional joint distribution function.

6.1 Bivariate digital options

A **digital option** is a contract that pays the underlying asset or one unit cash if the price of the underlying is above or below some strike level at maturity. Digital options paying a unit cash are called **cash-or-nothing (CoN) options**, while those paying the asset are called **asset-or-nothing (AoN) options**.

Consider two digital cash-or-nothing call options DC_1 and DC_2 with respective strikes K_1, K_2, written on the assets S_1, S_2 having copula C_{LL}. If at maturity $S_i > K_i$, then DC_i pays one unit cash, $i = 1, 2$. The prices of these digital call options are

$$\begin{aligned} DC_i(t,T,K_i) &= B(t,T)\mathbb{E}^{\mathbb{Q}}\left[\mathbf{1}(S_i(T) > K_i) \,|\, \mathcal{F}_t\right] \\ &= B(t,T)\mathbb{Q}(\{S_i(T) > K_i\} \,|\, \mathcal{F}_t), \end{aligned} \tag{22}$$

where $\mathbb{Q}$ is the risk-neutral measure associated with taking the discount bond $B(t,T)$, paying one unit cash at maturity, as the numéraire.

Analogously, for $i = 1, 2$, the prices of the digital cash-or-nothing put options DP_1 and DP_2 with strikes K_1, K_2, written on the assets S_1, S_2 are given by

$$DP_i(t,T,K_i) = B(t,T)\mathbb{Q}(\{S_i(T) < K_i\} \,|\, \mathcal{F}_t). \tag{23}$$

From (22) and (23) it follows that the value of a portfolio containing a digital call and a digital put equals the price of the discount bond.

Letting C_{HH} denote the survival copula to C_{LL}, the price of a **bivariate digital call** option paying one unit if both $S_1 > K_1$ and $S_2 > K_2$, is

$$\begin{aligned} D_{HH}(t,T,K_1,K_2) &= B(t,T)\, C_{HH}(\, \mathbb{Q}(\{S_1(T) > K_1\} \,|\, \mathcal{F}_t),\, \mathbb{Q}(\{S_2(T) > K_2\} \,|\, \mathcal{F}_t)\,) \\ &= B(t,T)\, C_{HH}\left(\frac{DC_1}{B}, \frac{DC_2}{B}\right). \end{aligned}$$

In the general bivariate case, one can no longer distinguish between 'put' or 'call'. Instead, subscripts will be added to describe the payoff. For example, H(igher) L(ower) means that one unit will be paid if the price of the first underlying is above a strike level and the second is beneath a (possibly different) strike level. The price of this option in fact follows from the sigma-additivity of $\mathbb{Q}$:

$$\begin{aligned} D_{HL} &= DC_1 - D_{HH} \\ &= DC_1 - B(t,T)C_{HH}\left(\frac{DC_1}{B}, \frac{DC_2}{B}\right) \\ &= B\left\{\frac{DC_1}{B} - C_{HH}\left(\frac{DC_1}{B}, \frac{DC_2}{B}\right)\right\} \\ &= B\, C_{HL}\left(\frac{DC_1}{B}, \frac{DP_2}{B}\right), \end{aligned}$$

where $C_{HL}(u,v) := u - C_{HH}(u, 1-v)$ satisfies the properties of a copula. In the last step the relation $DC_1 + DP_1 = B$ is used.

Using a similar argument one can show

$$D_{LH} = B\, C_{LH}\left(\frac{DP_1}{B}, \frac{DC_2}{B}\right),$$

with $C_{LH}(u,v) := v - C_{HH}(1-u, v)$ the survival copula to C_{HL}.

Finally, using the definition of survival copulas

$$\begin{aligned} \frac{D_{LL}}{B} &= 1 - \frac{DC_1}{B} - \frac{DC_2}{B} + C_{HH}\left(\frac{DC_1}{B}, \frac{DC_2}{B}\right) \\ &= C_{LL}\left(\frac{DP_1}{B}, \frac{DP_2}{B}\right). \end{aligned}$$

6.2 Rainbow options

Rainbow options, also known as 'best-of' options, are multivariate contingent claims whose underlying asset is the maximum or minimum in a set of assets.

Consider for example a put option on the maximum $Z(T) := \max\{S_1(T), S_2(T)\}$ of two assets:

$$\mathcal{X}(Z(T)) = \max\{0, K - Z(T)\}.$$

Integrating expression (38) over the interval [0, K] yields

$$PUT(Z,t;T,K) = e^{r(T-t)} \int_0^K \mathbb{Q}(Z(t) < u \,|\, \mathcal{F}_t)\, du.$$

For the maximum of the prices of two assets to be smaller than a certain value, both prices have to be smaller than that value:

$$PUT(S_1, S_2, t; T, K) = e^{r(T-t)} \int_0^K \mathbb{Q}(S_1(t) < u, S_2(t) < u \,|\, \mathcal{F}_t)\, du.$$

Replacing the joint distribution function by the copula C_{LL} from Section 6.1 gives

$$PUT(S_1, S_2, t; T, K) = e^{r(T-t)} \int_0^K C_{LL}(\mathbb{Q}(S_1(t) < u),\ \mathbb{Q}(S_2(t) < u) |\, \mathcal{F}_t)\, du.$$

7 Project plan

The goal of the project is to examine whether in modeling the dependence structure of contracts on tail dependent underlyings, the use of copulas different from the Gaussian copula affects the price.

Although the final goal is to price contracts on any number of underlyings, a starting point could be to fit a one-parameter bivariate copula to a given set of data using maximum likelihood. The margins will eventually have to be modeled parametrically so that volatility smile can be incorporated, but if we choose a two-step approach in the maximum likelihood estimation (IFM, section 5.2), the same estimation procedure can also be applied to empirical margins (i.e. CML, section 5.3). To test the appropriateness of the resulting model, the likelihood of observing the data in the new model can be compared to the case of a Gaussian copula. Another obvious check would be to compare sample versions of Kendalls tau, Spearmans rho and the TDCs to the corresponding quantities as implied by the fitted copula. Furthermore, the copula estimate should not change too much in time and also sensitivity to the parameters of the marginal distributions (SABR, displaced diffusion) has to be acceptable.

A next step could be to consider a mix of copulas, at least one of which has tail dependence. Maximum likelihood estimation now is more involved, since there are several parameters to fit. Hu (2004) suggests using the expectation maximization algorithm to carry out the second step in the IFM method.

The calibrated mix of copulas could then be implemented in a pricing model to see if the new copula indeed affects the price significantly.

Possible further steps include extension of the model to three or more dependent assets (since many products have more than two underlyings) and performing a hedge test.

A Basics of derivatives pricing

Consider a market model consisting of price processes

$$S(t) = \begin{pmatrix} S_1(t) \\ \vdots \\ \vdots \\ S_n(t) \end{pmatrix} \tag{24}$$

defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Also, let $B(t)$ be the money account. 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: what should be the price $\Pi(t; \mathcal{X})$ of the T-claim $\mathcal{X}$ at time t?

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 driven by the same Wiener process:

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

Construct the portfolio

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

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

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

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

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

or equivalently

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

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 [15]. 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:

$$\begin{aligned} dB(t) &= rB(t)dt, \\ dS(t) &= \mu S(t)dt + \sigma S(t)dW(t). \end{aligned} \tag{25}$$

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

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

$$\Pi(t) = F(t, S(t)) \tag{26}$$

where F is a smooth function. Applying Itō's formula to (26) and omitting arguments:

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

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

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

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

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

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 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 (28). 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*

$$\mathbb{E}^{\mathbb{P}}\left[e^{\frac{1}{2}\int_0^T ||\phi(t)||^2(t)\,dt}\right] < \infty, \tag{27}$$

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

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

For a proof, refer to Björk [16].

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 A.2 it follows that the process $S(t)/B(t)$ has zero drift. If the volatility of this process satifies the Novikov condition (27), then zero drift implies $S(t)/B(t)$ to be a martingale[1]. Pricing formula (28) is an immediate consequence of this.

A measure like $\mathbb{Q}$ 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 $\mathbb{Q}$ 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*

$$\frac{\Pi(t;\mathcal{X})}{S_0(t)} = \mathbb{E}^Q\left[\left.\frac{\Pi(T;\mathcal{X})}{S_0(T)}\right|\mathcal{F}_t\right] = \mathbb{E}^Q\left[\left.\frac{\mathcal{X}}{S_0(T)}\right|\mathcal{F}_t\right] \tag{28}$$

[1]In general, zero drift does not imply a stochastic process to be a martingale. The implication holds under an extra condition, see [17] p. 79. For exponential martingales, this condition is equivalent to the (more practical) Novikov condition.

where $\mathbb{Q}$ is the (not necessarily unique) martingale measure for the market $S_0, S_1, \dots S_n$ with S_0 as the numéraire.

Suppose you have a Wiener process to which you add a drift term. How do you have to change the measure to make the resulting Brownian motion a (driftless) Wiener process again under the altered measure? This question was answered by Girsanov's theorem.

Now suppose you have a martingale that you multiply by a (positive) stochastic process. The next lemma describes how you have to change the measure if you want the resulting process to be a martingale again under a new measure.

Lemma A.6 (Change of numéraire) *Assume that $\mathbb{Q}^0$ is a martingale measure for the numéraire S_0 (on $\mathcal{F}_T$) and assume that S_1 is a positive asset price process such that $S_1(t)/S_0(t)$ is a $\mathbb{Q}^0$ martingale. Define $\mathbb{Q}^1$ on $\mathcal{F}_T$ by the likelihood process*

$$L_0^1(t) = \frac{S_0(0)}{S_1(0)} \frac{S_1(t)}{S_0(t)}, \quad 0 \le t \le T. \tag{29}$$

Then $\mathbb{Q}^1$ is a martingale measure for the numéraire S_1.

Proofs of theorems A.4, A.5 and lemma A.6 can be found in Björk [16].

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

$$dS_i(t) = \alpha_i(t)S_i(t)dt + S_i(t)\sigma_i(t)dW^{\mathbb{P}}, \quad i = 0, 1\,,$$

Itō's formula applied to (29) gives the Girsanov kernel for the transition from $\mathbb{Q}^0$ to $\mathbb{Q}^1$:

$$\phi_0^1(t) = \sigma_1(t) - \sigma_0(t).$$

A **zero-coupon bond** is an asset that pays one unit currency at maturity T. The price $p(t,T)$ of such a bond at time t thus reflects the amount needed to ensure one unit currency in the future.

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** $\mathbb{Q}^T$.*

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

$$L_{\mathbb{Q}}^T = \frac{S_0(0)}{p(0,T)} \frac{p(s,T)}{S_0(s)} \tag{30}$$

relating $\mathbb{Q}$ to $\mathbb{Q}^T$. It follows that

$$\begin{aligned}
\frac{\Pi(s)}{p(s,T)} &= \frac{S_0(s)}{p(s,T)} \frac{\Pi(s)}{S_0(s)} \overset{\text{thm. A.5}}{=} \frac{S_0(s)}{p(s,T)} \mathbb{E}^{\mathbb{Q}} \left[\frac{\Pi(t)}{S_0(t)} \middle| \mathcal{F}_s \right] \\
&= \frac{\frac{S_0(0)}{p(0,T)} \mathbb{E}^{\mathbb{Q}} \left[\frac{\Pi(t)}{S_0(t)} \middle| \mathcal{F}_s \right]}{\frac{p(s,T)}{p(0,T)} \frac{S_0(0)}{S_0(s)}} \overset{(30)}{=} \frac{\mathbb{E}^{\mathbb{Q}} \left[L_{\mathbb{Q}}^T \frac{\Pi(t)}{p(t,T)} \middle| \mathcal{F}_s \right]}{L_{\mathbb{Q}}^T} \\
&\overset{\text{Bayes' form.}}{=} \mathbb{E}^T \left[\frac{\Pi(t)}{p(t,T)} \middle| \mathcal{F}_s \right],
\end{aligned}$$

where $\mathbb{E}^T$ denotes integration w.r.t. $\mathbb{Q}^T$. In particular, as $p(T,T) = 1$:

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

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

Lemma A.10 *$\mathbb{Q}^T$ is equal to the risk-neutral measure associated with the fixed money account iff. the interest rate r is deterministic.*

Proof The two measures $\mathbb{Q}^T$ and $\mathbb{Q}$ being equal implies their Radon-Nikodym derivative to be one. From equation (30) it can be seen that this is equivalent with the relation $p(t,T) = p(0,T)S_0(t)$ to hold for all $0 \leq t \leq T$. 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 $p(t,T)/S_0(t) \neq p(0,T)$ for some t clearly leads to arbitrage opportunities.

Lemma A.11 (Geman–El Karoui–Rochet) *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 $\mathcal{X} = \max\{0, S(T)-K\}$. The option price is*

$$\Pi(0;\mathcal{X}) = S(0)\mathbb{Q}^S(S(T) \geq K) - Kp(0,T)\mathbb{Q}^T(S(T) \geq K). \tag{31}$$

*Here $\mathbb{Q}^T$ denotes the T-forward measure and $\mathbb{Q}^S$ is the martingale measure for the numéraire process $S(t)$. Under the assumption that the process $\frac{S(t)}{p(t,T)}$ has **deterministic volatility** $\sigma_{S,T}(t)$, equation (31) reduces to*

$$\Pi(0;\mathcal{X}) = S(0)N[d_1] - Kp(0,T)N[d_2], \tag{32}$$

where

$$\begin{aligned} d_1 &= d_2 + \sqrt{\Sigma^2_{S,T}(T)}, && (33)\\ d_2 &= \frac{\log(\frac{S(0)}{Kp(0,T)}) - \frac{1}{2}\Sigma^2_{S,T}(T)}{\sqrt{\Sigma^2_{S,T}(T)}}, && (34)\\ \Sigma^2_{S,T}(T) &= \int_0^T ||\sigma_{S,T}(t)||^2 dt. && (35)\\ & && (36) \end{aligned}$$

B Implied quantities

A market model and observed data (such as stock prices, option prices) together form an overcomplete model. By removing the assumptions on some of the input parameters, the new model can be used to retrieve the 'implied' quantities, that is, implied by the market data *and* by the assumptions in the model that were not removed.

This idea can for instance be used to check whether assumptions on input parameters are reasonable.

B.1 Implied distribution

How are the prices of the assets in the market distributed? The distribution cannot be observed directly from market data, but it can be reconstruced from actual prices under specific assumptions. These assumptions entail a certain market model to hold — except maybe for the distribution of the prices — such that, together with the observed data, the distribution is implied by the model. The implied distribution is sometimes referred to as the 'implied measure'.

As an example, consider the Black-Scholes model of Example A.3. The solution of the SDE (25) 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, i.e. the probability of extreme values of the stock, are thicker than assumed in Black-Scholes.

Breeden and Litzenberger [18] proposed a way to reconstruct what would be the distribution, if lognormality of the returns does not hold, but all the other assumptions of Black-Scholes are satisfied. Denote $CALL(S,t;T,K)$ the price of a call option with maturity T and strike K on an asset S. Let $\mathbb{Q}$ be the EMM associated with the money account and let $\mathbb{Q}^S$ be the martingale measure for the price process of the call option with numéraire $S(t)$. From the Geman–El Karoui-Rochet lemma (A.11), assuming furthermore constant volatility ($\Sigma^2_{S,T} = \sigma^2$) and constant interest rate ($\mathbb{Q}^T = \mathbb{Q}$, lemma A.10), we have:

$$\begin{aligned} CALL(S,t;T,K) &= S(t)\mathbb{Q}^S(S(T) \geq K \,|\, \mathcal{F}_t) - Ke^{-r(T-t)}\mathbb{Q}(S(T) \geq K \,|\, \mathcal{F}_t) \\ &= S(t)\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2). \end{aligned}$$

Differentiation with respect to K yields, after some calculation,

$$-e^{r(T-t)}\frac{\partial CALL(S,t;T,K)}{\partial K} = \mathbb{Q}(S(T) \geq K \,|\, \mathcal{F}_t). \tag{37}$$

Similarly, for the price $PUT(S,t;T,K)$ of a put option with underlying S, maturity T and strike K it can be shown that

$$e^{r(T-t)}\frac{\partial PUT(S,t;T,K)}{\partial K} = \mathbb{Q}(S(T) \leq K \,|\, \mathcal{F}_t). \tag{38}$$

B.2 Implied volatility

One can also assume the Black-Scholes model to be correct, except for the constant volatility. Assuming thus the pricing formula for a call option to be correct, one could for instance find the volatility for which the Black-Scholes price coincides with the market price. Doing so for different values of the strike yields the so called 'volatility smile', i.e. the effect that in practise volatility is not constant, but relatively higher for extreme strike prices. Calculating the implied volatility for different strike–maturity pairs gives what is know as the 'volatility surface'.

Dupire [19] showed that under risk neutrality, there is a unique **local volatility function** $\sigma(t, T)$ consistent with the implied distribution from the previous section. The implied volatility, however, is much easier to infer from the market and it also facilitates the handling of time varying distributions.

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

[13] 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.

[14] A.J. Patton. Modelling time-varying exchange rate dependence using the conditional copula. Discussion paper, University of California, 2001.

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

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

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

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

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