Journal of Mathematics and Statistics

Original Research Paper

Multivariate Option Pricing with Gaussian Mixture Distributions and Mixed Copulas

[1]Jimbo Henri Claver, [2]Tatanfack Emerson and [2]Shu Felix

[1]Department of Mathematics and Computing, American International University, Kuwait
[2]Department of Mathematics and Computer Science, University of Bamenda, Cameroon

Article history
Received: 19-09-2022
Revised: 19-12-2022
Accepted: 24-12-2022

Corresponding Author:
Jimbo Henri Claver
Department of Mathematics and Computing, American International University, Kuwait
E-mail: jimbo.maths@gmail.com

Abstract: Recently, it has been reported that the hypothesis proposed by the classical black Scholes model to price multivariate options in finance were unrealistic, as such, several other methods have been introduced over the last decades including the copulas methods which uses copulas functions to model the dependence structure of underlying assets. However, the previous work did not take into account the use of mixed copulas to assess the underlying assets' dependence structure. The approach we propose consists of selecting the appropriate mixed copula's structure which captures as much information as possible about the asset's dependence structure and apply a copulas-based martingale strategy to price multivariate equity options using monte Carlo simulation. A mixture of normal distributions estimated with the standard EM algorithm is also considered for modeling the marginal distribution of financial asset returns. Moreover, the Monte Carlo simulation is performed to compute the values of exotic and up and out barrier options such as worst of, spread, and rainbow options, which shows that the clayton gumble and clayton gaussian have relatively large values for all the options. Our results further indicate that the mixed copula-based approach can be used efficiently to capture heterogeneous dependence structure existing in multivariate assets, price exotic options and generalize the existing results.

Keywords: Monte Carlo Simulation, Dependence Structure, Exotic and Barrier Options, Copulas Method, Gaussian Mixture Distributions, and Mixted Copulas, Black Scholes Model

Introduction

The pricing of derivatives using copulas has received extensive attention within the actuarial and finance literature in the past two decades. For instance, Chiou and Tsay (2008) priced multivariate exotic derivatives using copula-based models and the risk neutral representation and demonstrated how copula-based models may be used to determine the value at risk of numerous assets. Kim and Kim (2015) used copula functions to derive correlation coefficients that depend on strike prices between assets rather than the more straightforward correlation coefficients to offer an accurate technique of pricing rainbow options with stochastic simulation. Malgrat (2013) used the copula method to model the underlying dependency structure of assets and price basket options by employing several families of copulas with parameters chosen using the maximum likelihood approach. Church (2011) used a similar idea and devised a copula-based algorithm for pricing path dependent basket options and demonstrated how the marginals selected can significantly affect the price of the options. Barban and Di Persio (2014); Berton and Mercuri (2021) presented an alternative to the Monte Carlo simulation method to price multivariate options by using a copula GARCH model. Hassane *et al.* (2021) presented an approach to price options that provide for the modeling of financial asset returns to take the impact of extreme values into consideration by using a combination of two gaussian distributions and an extreme copula to model the returns' combined dependence structure.

The Multivariate Geometric Brownian Motion (MGBM) approach is the most commonly used in the literature and in practice for pricing options. This approach requires an unrealistic hypothesis of independence or perfect correlation between the underlying assets. During the last decade, several alternatives have been proposed to solve this problem. The stochastic volatility and copulas methods are two examples. The stochastic volatility approach models volatility as a stochastic quantity. One of the most widely used models which assume nonconstant volatility of assets was proposed by Heston (1993). For a better understanding

DOI: 10.3844/jmssp.2023.1.12

of the Heston model, (Ball and Roma, 1994; Kouritzin and Mackay, 2020). The copulas method's main goal is to distinguish the marginal distributions of variables from their reliance on one another. Flexible dependence structures are possible, such as tail dependent or nonlinear dependence (Chiou and Tsay, 2008). Another unrealistic assumption that has been abandoned over time is the joint normality of asset returns, which does not account for distribution tails or extreme events. In this case, the use of mixtures of gaussian distributions has been shown to be preferable (Hassane *et al*., 2021) due to the fact that it approximates all the distributions previously used (gaussian, student, hyperbolic, etc.).

Multivariate option pricing using copulas is a field in finance that has been developing very fast for the last two decades. In many works on the pricing of options with copulas, the joint risk neutral density is considered after marginal risk neutral densities or risk neutral copula (Hassane *et al*., 2021). However, all this past work did not take into account the use of mixed copulas to assess the underlying assets' dependence structure. The ability of the mixed copula model to nest various copula shapes is by far its greatest benefit. Therefore, compared to an individual copula, a mixed copula is empirically more adaptable in modeling dependence structure and can provide better descriptions of dependence structure (Cai and Wang, 2014).

In the present work, we suggest an approach for multivariate options valuation, allowing the dependent structure to be written as a copula mixture and analyzing how our choice of copula affects the cost of multivariate exotic and barrier choices with a range of strikes. In addition, we price exotic options using a copulas-based method by describing the statistical behavior of each underlying asset using a mixture of gaussian distributions, which have been proven to capture excess kurtosis and asymmetry. In particular, we model the underlying asset returns as a mixture of two gaussian distributions, with two components: One capturing the more volatile "crisis" behavior and the other the "business as usual" stock return behavior with a mean close to zero. Also, the modeling of the dependencies between assets using different mixtures of copulas distributions so as to include a variety of dependency structures.

This study is set up as follows: The main copula principles are briefly reviewed in section two. In the third section, we expose the methodology for the martingale option pricing approach with mixed copulas. Then, in the fourth section, the results of different estimations and simulations are presented, along with their analyses and interpretations. The last section presents a conclusion and discussion.

Materials and Methods

Facts About Copulas

In this section, we recall the basic notion, definition, and theorem about copulas. For more details on copulas, you may refer to Pfaff (2016); Hofert *et al*. (2018). Copulas are mathematical functions that represent the relationships (dependency) among random variables.

Definition 1.1

N-dimensional copulas are multivariate distribution functions. (d.f.) C, with margins uniformly distributed in [0,1] (U(0,1)) and the following properties:

1. $C: [0,1]^n \to [0,1]$;
2. C is grounded and increasing
3. C has margins C_i which satisfy $C_i = C(1,...,1,u,1,...,1) = u$ for all $u \in [0,1]$

The main outcome of the copula theory is the well-known Sklar theorem by Sklar (1959). As stated therein.

Theorem 1.2

Sklar theorem for any real valued random variables $X_1,..., X_n$ with joint distribution function F and univariate marginal distribution functions $F_1,..., F_n$ there exists a copula C such that:

$$F\left(x_1,...,x_n\right) = C\left(F_1\left(x_1\right),...F_n\left(x_n\right)\right) \quad (1)$$

In contrast, given a copula C and univariate distribution functions $F_1,..., F_n$, then the function F defined by (1) is a d-dimensional distribution function with univariate margins $F_1,..., F_n$.

There are many proofs of this theorem that arise from different properties of copulas.

A proof by multilinear interpolation was established, first in the bivariate situation, by Schweizer and Sklar (1974), and later in the multivariate case by Carley and Taylor (2002). Moore and Spruill (1975) used the idea of generalized probability integral transformation to prove the theorem; (Durante *et al*., 2012) showed the existence of a random vector's connection to a copula without presenting analytically or probabilistically its form; and finally, (Benth *et al*., 2022) have demonstrated a topological demonstration of the theorem in any dimension.

Furthermore, the copula is unique if $F_1,...,$ and F_n are continuous functions. Sklar's theorem gives a flexible approach to the construction of multivariate distributions and to derive from any monotonic transformation certain features of the copula, such as invariance, which is particularly useful in financial applications where logarithmic conversions are commonly utilized. Also, every copula C satisfies:

$$\max\left(\sum_{i=1}^{d} u_i - n + 1; 0\right) \le C\left(u_1, u_2, ..., u_n\right) \le \min\left(u_1, u_2, ..., u_n\right)$$

where the upper bound for all $d = 2$ is a copula and the lower bound for all $d \geq 2$ is a copula.

DOI: 10.3844/jmssp.2023.1.12

Classes and Families of Copulas in Finance

Copulas come in a variety of families, some of which are more suitable for financial modeling; these families are stated in the next few subsections.

Elliptical Copulas

One of the most popular types of copulas in use today is the elliptical one. With the aid of Sklar's theorem, they are created from elliptical distributions and are used to dependence structure these elliptical distributions. The normal and *t*-copulas are the two most prominent elliptical copulas and they have the following respective density functions:

$$C\left(u_1,...,u_n;\Sigma\right) = \frac{1}{\|\Sigma\|^{1/2}}\exp\left[-\frac{1}{2}\xi\left(\sum{}^{-1}-l\right)\xi\right] \tag{2}$$

where, $\xi = \left(\Phi^{-1}\left(u_1\right),...,\Phi^{-1}\left(u_n\right)\right)$ with Φ^{-1} the quantile function of the univariate standard normal distribution and l the d-dimensional identity matrix:

$$C\left(u_1,...,u_n;v,\Sigma\right) = \frac{\Gamma\frac{\left(v+n\right)}{2}\left[\Gamma\left(\frac{v}{2}\right)\right]^{n-1}\|\Sigma\|^{-1/2}}{\left[\Gamma\left(\frac{v+1}{2}\right)\right]^{d}} \times \frac{\left(1+\frac{1}{v}\xi\Sigma^{-1}\xi\right)^{-\frac{v+n}{2}}}{\prod_{i=1}^{n}\left(1+\frac{\xi_i^2}{v}\right)^{-\frac{v+1}{2}}} \tag{3}$$

$\xi = \left(t_v^{-1}\left(u_1\right),...,t_v^{-1}\left(u_n\right)\right)$ compared to normal copulas, the *t*-copulas are more efficient at modeling tail dependencies in the distribution with an additional parameter, the degrees of freedom v. As n gets smaller, the dependence gets stronger. The *t*-copula converges to the normal copula for $s \to \infty$.

Archimedean Copulas

The Archimedean copulas, which may be easily built with univariate generating functions and have closed forms, represent another significant family of copulas. The advantage of these copulas is that they may capture the pattern of positive or negative dependence between the variables (asymmetrical dependence) patterns). The main disadvantage is their small number of parameters, which reduces flexibility. An archimedean copula is a copula of the form:

$$C\left(u_1,...,u_n\right) = \psi\left(\psi^{-1}\left(u_1\right)+...+\psi^{-1}\left(u_n\right)\right) \tag{4}$$

where, $u_1,..., u_n \in [0,1]$ for a so called generator ψ: $[0,\infty] \to [0,1]$ which satisfies $\psi(0) = 1$, $\psi(\infty) = lim_{t\to\infty}(\psi(t)) = 0$ and which is strictly decreasing on $[0,\inf\{t: \psi(t) = 0\}]$. It follows from (4) that archimedean copulas are exchangeable.

The density of the Archimedean copulas is given by:

$$C\left(u_1,...,u_n\right) = \psi^{(n)}\left(\psi^{-1}\left(u_1+...+\psi^{-1}\left(u_n\right)\right)\right)\prod_{i=1}^{n}\psi\left(u_i\right) \tag{5}$$

where, $\psi^{(n)}$ is the n^{th} mixed partial derivative of the inverse generator.

The clayton, frank, and gumbel copulas are among the members of the family of archimedean copulas, which is primarily employed in the area of finance. Table 1 their respective generator functions and distributions.

where, $\theta \geq 0$ for clayton and frank and $\theta \geq 1$ for gumble.

An increase in the parameter theta suggests a stronger dependence for each of these archimedean copulas.

Estimation and Conformity Test of Copulas

Considering the marginal distributions of $F_1,...,$ and F_n is unknown, the copulas C need to be estimated assuming that they are members of a parametric, semiparametric, or nonparametric copula family.

Estimation

If we consider $F_1,..., F_n$ to be members of absolutely continuous parametric families of univariate distributions. Assuming that the unknown copula C_{θ_0} belongs to the parametric family of distribution functions, the estimation problem of interest entails estimating it. The Inference Functions for Margin Estimator (IFME), a two-stage estimator, is used to achieve this:

1. First, the parameters α_i, $i = 1,...,n$ of the marginals distributions are estimated via maximum likelihood:

$$\hat{\alpha} = argmax\sum_{j=1}^{N}\ln f_d\left(x_i^j;\alpha_i\right) \tag{6}$$

2. Next, the copula's parameter, θ that, according to maximum likelihood estimation, best describes these marginals:

$$\hat{\theta} = argmax\sum_{j=1}^{N}\ln C\left(F_1\left(x_i^j;...;F_n\left(x_i^j;\theta\right)\right)\right) \tag{7}$$

Because it requires less processing, this IFME estimation approach is more effective than the accurate MLE method.

Conformity Test

This is a fit test that fits chosen copula model to the data well. The most effective tests are based on the processes $\sqrt{n}\left(\hat{C}-C_{\hat{\theta}}\right)$ where $\hat{C}$ and $C_{\hat{\theta}}$ are, respectively, the empirical copula and the parametric copula.

The Akaike Criterion (AIC) is used frequently to select the best copula:

$$AIC = 2m - \log\left(l(\theta)\right) \quad (8)$$

where, m is the number of parameters to be estimated, n is the size of data, and $l(\theta)$ is the model likelihood for the estimated parameter θ.

Copulas Mixtures

Several different copula groups are combined linearly to form a mixed copula. A mixed copula function is defined mathematically as:

$$mix\left(C_1,...,C_m\right)(u,\theta) = \sum_{k=1}^{m}\lambda_k C_k(u,\theta)$$
$$= \sum_{k=1}^{m}\lambda_k C_k\left(F_1\left(x_1,\alpha_1\right),...,F_p\left(x_p,\alpha_p\right);\theta_k\right)$$

where, $C_1,..., C_m$ is a collection of known base copulas with unidentified parameters. $\{\theta_k\},\{\lambda_k\}_{k=1}^{m}$ are the priori probabilities satisfying $0 \leq \lambda_k \leq 1$ and $\sum_{k=1}^{m}\lambda_k = 1$ and m is the number of components for all $k \in \{1,..., m\}$. When only one component is present in a mixed copula, it is known as a single copula. All linear functionals of the mix $(C_1,..., C_m)$ (u, θ) can be calculated as the equivalent mixture of linear functionals of the component $C_1,..., C_m$. For instance, in the bivariate case, Spearman's rho, the coefficient of lower tail dependence, the coefficient of upper tail dependence, and the density of mix $(C_1,..., C_m)(u, \theta)$ are simply:

$$\sum_{k=1}^{m}\lambda_k \rho_m\left(C_k\right), \sum_{k=1}^{m}\lambda_k w_l\left(C_k\right), \sum_{k=1}^{m}\lambda_k w_u\left(C_k\right) and \sum_{k=1}^{m}\lambda_k C_k(u)$$

Note that sampling from a mixture of copulas is immediate from the sampling of individual components.

The copulas used in finance are mostly from the archimedean and elliptical families of copulas, although the survival copulas are also used. Each of these copulas has its own characteristics that help it model the dependency structure. For example, gaussian copulas better model the linear dependence but are symmetrical, which is not convenient for fat tailed distributions; on the other hand, gumble copulas are asymmetrical and are best for right tail distributions. Hence, mixing these copulas helps to combine their different properties into a single copula.

Fitting Mixtures to Data

The three most commonly used methods are pseudo maximum likelihood, maximum likelihood, and Empirical Cumulative Density Function (ECDF).

A two-step process is used in the pseudo maximum likelihood:

- Map each marginal piece of data to its quantile using the ECDF
- For the quantile data, determine Kendall's $\hat{\tau}$ and use it to determine $\hat{\theta}$

For elliptical copulas, θ estimated here is their ρ, the correlation parameter.

As for the maximum likelihood and ECDF, the model and likelihood function can be optimized from copula density $c(u_1, u_2)$ as the objective function to maximize. This is slow, though, and even for a bivariate copula with arbitrary univariate marginal distributions, it might not be the best solution with many parameters to fit. Hence, in general, we don't use this strategy and ECDFs are instead frequently used.

Table 1: Generating function, distribution, tail dependence, and tau of Archimedean copulas

Copulas	ψ_θ^{-1}	$C_\theta(u_1,...,u_n)$	λ_L	λ_U	τ
Clayton	$\frac{1}{\theta}\left(t^{-\theta}-1\right), \theta \geq 0$	$\left[\left(\sum_{i=1}^{n}u_i^{-\theta}\right)+1\right]^{-\frac{1}{\theta}}$	$2^{\frac{-1}{\theta}}$	0	$\frac{\theta}{\theta+2}$
Gumble	$(-\ln t)^{\theta}, \theta \geq 1$	$\exp\left(-\left[\sum_{i=1}^{n}\left(-\ln u_i\right)^{\theta}\right]^{\frac{1}{\theta}}\right)$	0	$2-2\frac{1}{\theta}$	$1-\frac{1}{\theta}$
Franck	$-\ln\frac{e^{-\theta t}-1}{e^{-\theta}-1}\theta \geq 0$	$\frac{1}{\theta}\ln\left(1+\frac{\prod_{i=1}^{n}\left(e^{-\theta u_i}-1\right)}{\left(e^{-\theta}-1\right)^{n-1}}\right)$	0	0	$1-\frac{4}{\theta}+\frac{4}{\theta^2}\int_0^{\theta}\frac{t}{e^t-1}$

Valuation Approach for Multivariate Options

The objective is to price options whose underlying assets are risky assets $(S_t)_{t\in\mathbb{R}_+}$, where:

$$dS_t = \mu_t S_t dt + \sigma_t dW_t \tag{9}$$

where, μ_t and σ_t are constants and W_t is a standard geometric Brownian motion.

A well-known result from the martingale option pricing theory is the fundamental theorem of option pricing.

Theorem 2.1

Let $(\Omega, \mathscr{F}, (\mathscr{F})_{0\le t\le T}, \mathbb{P})$ be a filtered probability space. In No Arbitrage Opportunity (NAO), there exists a risk neutral measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ under which $\left\{e^{-rt}S_t\right\}_{t=0}^{T}$ is a $\mathbb{Q}$-martingale and such that at any time $t \in [0,T]$, the value of the contingent claim payout (fair or without arbitrage) P is given by:

$$V_t = e^{-(T-t)r}\mathbb{E}\left[P \mid \mathscr{F}_t\right] = e^{-(T-t)r}\int Pf\left(X_T\right)dX_T \quad 0 \le t \le T \tag{10}$$

Where, $T \ge 0$ is the option's maturity time, r is a real positive parameter that usually represents the risk-free rate and $f(X_T)$ is the underlying assets' joint risk neutral density probability function.

Recall that a market is without arbitrage opportunity if and only if it admits at least one equivalent risk neutral probability measure $\mathbb{Q}$.

Proposition 2.2

The price at time $t \in [0,T]$ of European call and put options for complete markets with strike price K, risk free rate r, and maturity T are respectively given by:

$$\begin{aligned} V_t &= e^{-(T-t)r}\mathbb{E}\left[\left(K - S_T\right)^+ \mid \mathscr{F}_t\right] \\ &= Ke^{-(T-t)r}\Phi\left(-d_-(T-t)\right) - S_t\Phi\left(-d_+(T-t)\right) \end{aligned}$$

and:

$$\begin{aligned} V_t &= e^{-(T-t)r}\mathbb{E}\left[\left(S_T - K\right)^+ \mid \mathscr{F}_t\right] \\ &= S_t\Phi\left(-d_+(T-t)\right) - Ke^{-(T-t)r}\Phi\left(-d_-(T-t)\right) \end{aligned}$$

with:

$$d_+(T-t) = \frac{\log\left(S_t / K\right) + \left(r + \sigma^2/2\right)(T-t)}{\sigma\sqrt{T} - t} \tag{11}$$

$$d_-(T-t) = \frac{\log\left(S_t / K\right) + \left(r - \sigma^2/2\right)(T-t)}{\sigma\sqrt{T} - t} \tag{12}$$

where, Φ is the standard gaussian cumulative distribution function and $0 \le t \le T$.

The study attempts to uncover valuation implications resulting from the use of dependence structures in mixed copulas and mixed marginal distributions. The normality of stock returns is a fundamental assumption in mathematical finance. However, empirical evidence often contradicts this theoretical foundation. As a consequence, the application of alternative models is an important research topic in financial modeling. The limitations of this assumption are.

Stock prices are frequently assumed to follow a geometric Brownian motion process; therefore, financial returns are considered to be independent and normally distributed, but the geometric Brownian motion assumption has many shortcomings.

The theoretical distribution, thus estimated, guarantees the exact reproduction of the mean and the variance. On the other hand, some empirical characteristics of the distribution of returns are not reproduced by the gaussian law:

1. Compared to the tails of the gaussian law distribution, the empirical distribution's tails are thicker. Extreme values of the returns are more often observed, hence producing a higher kurtosis
2. Because the empirical distribution is asymmetric compared to the gaussian distribution, it has a negative skewness coefficient

Finite Gaussian Mixtures

A Gaussian Mixture (GM) distribution model is a multivariate distribution that consists of multivariate gaussian distribution components. The mixture is defined by a vector of mixing proportions and each component is specified by its mean and covariance. It has been shown, for example, by Cha (2010) that the gaussian mixture with two or more normal distributions increases the model fit dramatically and captures the shape of the empirical distribution.

A GM model represents a distribution as:

$$f(x) = \sum_{j=1}^{K}\pi_j\mathbb{N}\left(x \mid \mu_k, \Sigma_k\right), \sum_{j=1}^{K}\pi_j = 1 \; and \; \pi_j \ge 0 \forall j \tag{13}$$

where, $\mathbb{N}(x|\mu_k, \Sigma_k)$ is the density function of multivariate Gaussian distribution.

DOI: 10.3844/jmssp.2023.1.12

Also, π_j represents the prior probability for Ω_j and each $\frac{1}{\sigma_j}\phi\left(\frac{x-\mu_2}{\sigma_2}\right)$ is the conditional probability density of X given Ω_j. Hence for every realization x of X, the posterior probabilities for each regime are given by:

$$P\left[\Omega_j \mid X = x\right] = \frac{\frac{\pi_j}{\sigma_j}\phi\left(\frac{x-\mu_j}{\sigma_j}\right)}{\sum_{h=1}^{k}\frac{\pi_h}{\sigma_h}\phi\left(\frac{x-\mu_h}{\sigma_h}\right)}$$

Numerous methods, such as Bayesian approaches, the method of moments, maximum likelihood, and graphical methods, can be used to fit mixture distributions. Due to the existence of an underlying statistical theory, the maximum probability method rises to the top among the others.

The EM algorithm is the method of preference for estimating the mixing model's parameters since it produces simple estimators. Let x_1, x_2,..., x_n be an observed random sample coming from a univariate GM and let:

$$\theta = \left(\pi_1, \pi_2, ..., \pi_k, \mu_1, \mu_2, ..., \mu_k, \sigma_1^2, \sigma_2^2, ..., \sigma_k^2\right)$$

be its vector parameters, then the corresponding likelihood function is:

$$l(\theta) = \sum_{i=1}^{n}\ln\left(\sum_{j=1}^{k}\frac{\pi_j}{\sigma_j}\phi\left(\frac{x-\mu_j}{\sigma_j}\right)\right)$$

in order to estimate θ, l must be maximized subject to $\sum_{j=1}^{k}\pi_k = 1$.

The EM Algorithm

The maximum likelihood estimate of θ can be obtained by the following iterative process known as the EM algorithm:

Algorithm 1: EM algorithm

Input: Data τ, initial guess $\theta^{(0)}$, and evaluate the log likelihood with these parameters

Output: Approximation of the maximum likelihood estimate.

1 t ← 1 **while** *change in Log-likelihood* ≥ ε **do**

Expectation Step: Evaluate the posterior probabilities $P^{(t)}(Z_i|\tau)$ with $\theta^{(t-1)}$ and compute the expectation

2 $Q^t(\theta) := \mathbb{E}_{P(t)}\left[\ln P\left(Z, \tau \mid \theta^{(t)}\right)\right]$

Maximization Step: Let $\theta^{(t)} \leftarrow \text{argmax}_{\theta \in \Theta} Q^t(\theta)$

3 $t \leftarrow t+1$

Return: $\theta^{(t)}$

For more information on the fitting of mixture distributions and the EM algorithm, (Ghojogh *et al*., 2019).

Several families of distributions have been proposed to overcome the shortcomings of the gaussian law:

- Alpha stable distributions
- The finite mixtures of distributions, such as gaussian mixtures
- Simple and generalized student's *t*-distributions
- Hyperbolic distributions

We are interested here in the mixture of gaussian distributions for the following reasons:

- It is simple to simulate
- It allows us to correctly approach all the alternative distributions cited above
- It makes it possible to reproduce various features observed in the data such as average, kurtosis, skewness, and variance
- It has several theoretical characteristics that make it simple to manipulate in the frame of a hypothetical asset price valuation model

Results and Discussion

In our application, we work with the dataset downloaded from the Yahoo finance website. In particular, we consider the following returns: BNP Paribas SA (BNP.PA) and Louis Vuitton stocks (LVMUY) Fig. 2. All daily data is extracted from Yahoo finance for a total of 237 days, from January 1, 2020, to December 31, 2020 (Fig. 1), resulting in the same number of closing levels for each return. We write an option contract X based on this data, with the payoffs P in Table 2 at maturity $T \geq 0$:

First, we correct the database with respect to time synchronization between closed prices and convert them to logarithmic returns (log returns). For each asset, Eq. (14) converts a closed price series S into a log returns r series:

$$r_{t,i} = \log\frac{S_{t,i}}{S_{t-1,i}} \tag{14}$$

where, $i \in (1,d)$ is the number of assets, $t \in (1,T)$ is a time point, in our case $d = 2$, $T = 237$ where $K \geq 0$ is the strike price of the option X and $B \geq 0$ is the barrier level our approach consists of the following steps:

1. Generate log return data from our historical data returns following Eq. (14)

2. Estimate the parameters of the gaussian mixture regime
3. Transform the log return sample into the pseudo sample
4. Estimate the parameters of the copulas mixtures using copulas suitable for financial applications
5. Calculate the option prices using the monte Carlo numerical integration method

Visualizing the above graphs, the following observations are made about the stock price for both BNP and LVMUY:

- Stock prices never reach the zero value
- Stock prices are continuous everywhere
- Stock prices are never negative
- Stock prices show randomness
- stock prices increase in the long run

From the quantile plots (Figs. 3-4) of the returns and the value of the kurtosis (greater than 3) in the descriptive statistics in Table 3, we can infer that the returns are not normally distributed.

At first, using the Expectation Maximization (EM) algorithm, the parameters of the two gaussian regimes that make up the gaussian mixture are calculated for each of the two assets. We shall limit ourselves to a mixture of two gaussian distributions. Also, we underline the presence of a fat tail for the returns of returns, in agreement with the non-gaussianity of the results.

Table 2: Options payoff

Option	Payoff
Worst option	$X(T) = \max\left(\min\left\{\frac{S_1(T)}{S_1(0)}, \frac{S_2(T)}{S_2(0)}\right\} - K, 0\right)$
Spread option	$X(T) = \max\{S_1 - S_2 - K; 0\}$
Up and out barrier	$X(T) = \max(S_1(T) + S_2(T)\text{-}k)1_{S_1(t) + S_2(t) \leq B}$

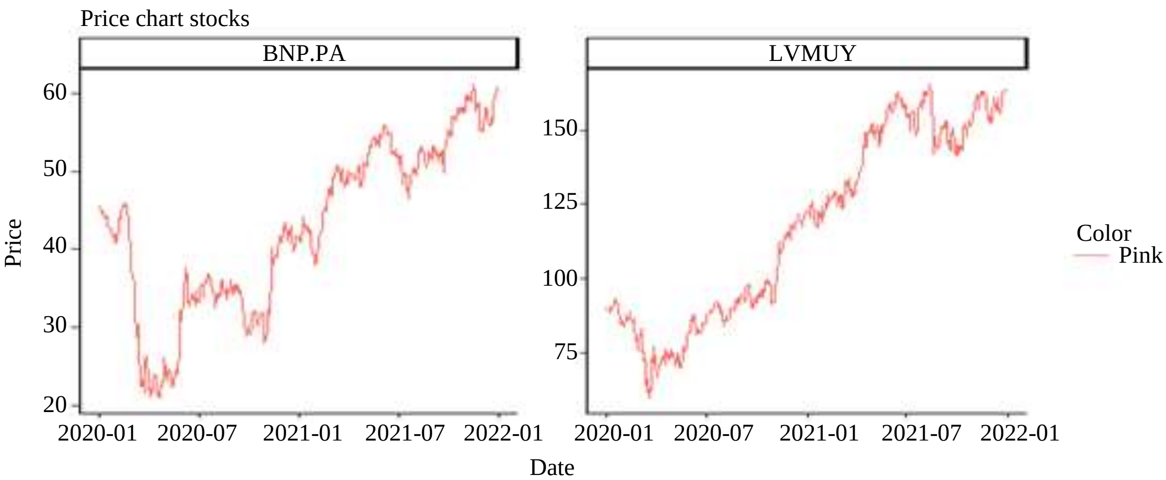

Fig. 1: Stocks dynamic

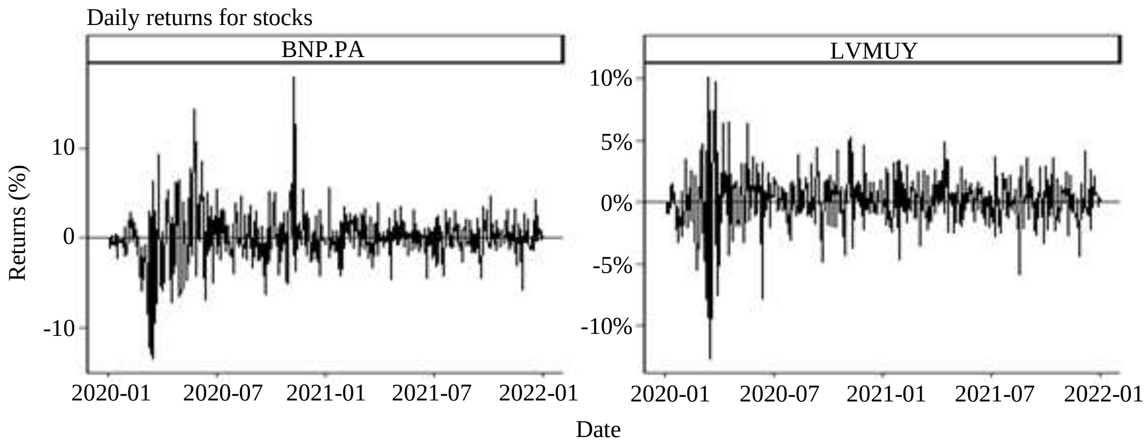

Fig. 2: Stocks return

DOI: 10.3844/jmssp.2023.1.12

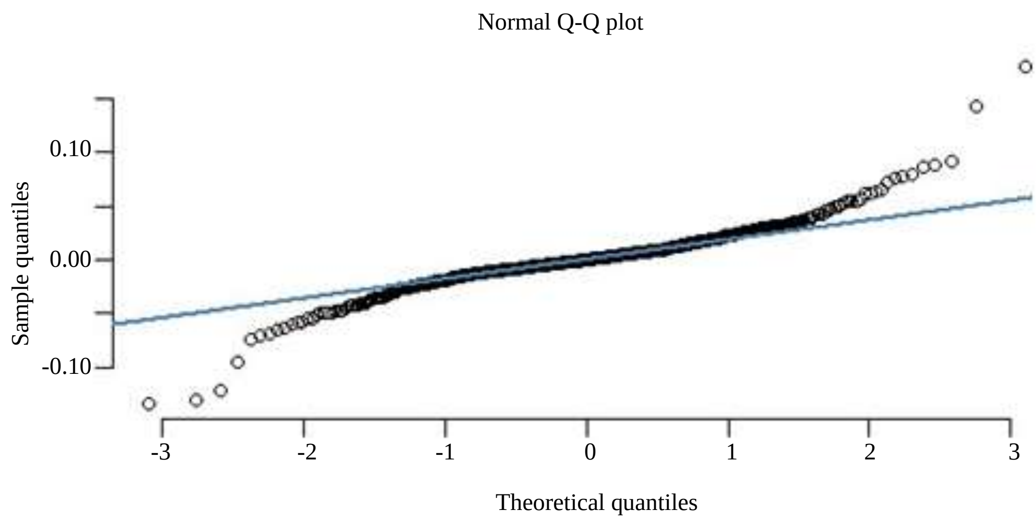

Fig. 3: QQ-BNP

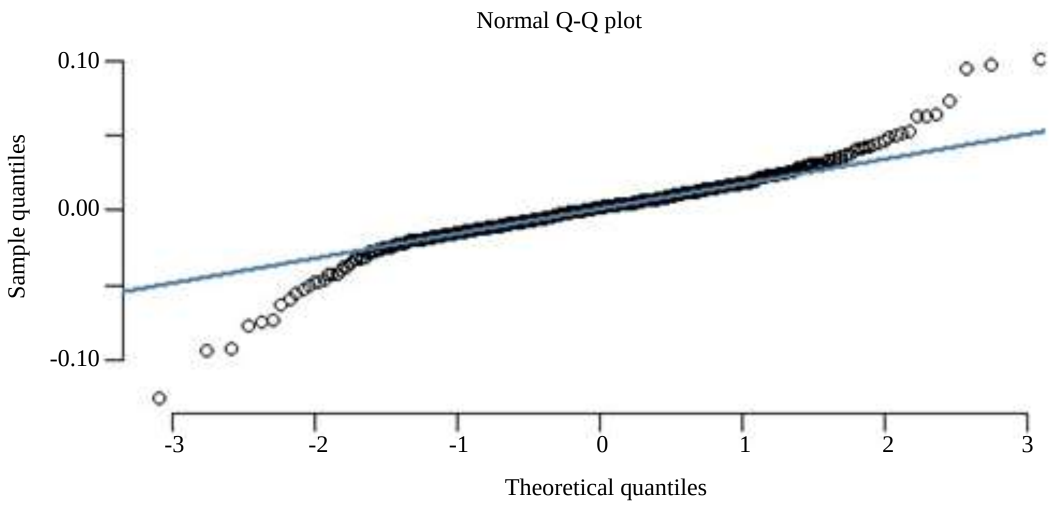

Fig. 4: QQ-LVMUY

Table 3: Gaussian mixture parameters

		Regime 1	Regime 2	Gaussian mixture	Empirical distribution
BNP.PA	Mean	0.0002059686	-0.003024669	-0.0006774372	-0.00046660
	Variance	0.0254138500	0.063095280	0.0013487840	0.02143490
	Skewness	0.0000000000	0.000000000	-0.1245244000	-0.18622490
	Kurtosis	3.0000000000	3.000000000	6.0715980000	3.25609200
	Proportion	0.7918247000	0.208175300	-0.0000000000	-0.00000000
	Mean	0.0003270575	-0.083608300	0.0016211340	0.00149100
LVMUY	variance	0.0229391500	0.026349790	0.0007847089	0.01630201
	Skewness	0.0000000000	0.000000000	0.3588690000	-0.61618140
	Kurtosis	3.0000000000	3.000000000	3.7702700000	4.29672600
	Proportion	0.9658321000	0.034167900	-0.0000000000	-0.00000000

DOI: 10.3844/jmssp.2023.1.12

Table 4: Copulas mixture parameters

Mixture	Values	$\hat{\sigma}$	$\hat{\theta}$	$\hat{\pi}$ 1	(CLTD, CUTD)
	Estimated value	-0.064256780	1.88077672	0.90251515	(0,0.05404298)
Clayton gumble	Standard error	0.088591830	1.27086298	0.13036934	
	Estimated value	-0.023207010	1.92443290	0.92202054	(0,0.04416826)
Gaussian gumble	Standard error	0.099159210	1.55531956	0.13535726	
	Estimated value	0.118835900	-0.21130660	0.59644730	(0,0)
Gaussian clayton	Standard error	0.135612180	0.09436311	0.32229041	
	Estimated value	-0.052031380	0.36477681	0.19598048	(0,0)
Gaussian frank	Standard error	0.438890700	1.48452170	2.10518570	
	Estimated value	-0.115452400	1.05600050	0.54800890	(0,0)
Clayton franck	Standard error	0.175625600	2.25166300	0.81296650	
	Estimated value	2.036135240	0.02156605	0.05678242	(0, 0.03375477)
Gumble franck	Standard error	0.175625600	2.25166300	0.81296650	

Table 5: Fitted Copulas mixture parameters

Mixture	$\hat{\sigma}$	$\hat{\theta}$	$\hat{\pi}$ 1	Statistic of test
Clayton gumble	-0.2608	1.5703	0.6011	3.692
Gaussian gumble	-0.6325	1.2304	0.2788	8.889
Gaussian clayton	-0.5161	0.9985	0.4266	12.07
Gaussian frank	-0.5739	4.0333	0.4314	13.33
Clayton franck	-0.5127	0.7306	0.1924	5.542
Gumble franck	1.0000	7.5821	0.8117	3.307

In the next step, we determine the parameters of the copula's mixtures, which are shown in Table 4. The estimation of the copula parameters and weights is jointly obtained by the minimization of the negative log likelihood of the weighted densities from the copulas, as in Pfaff (2016). Copula densities are computed as in Hofert *et al*. (2018). Four commonly used copulas in the finance field, namely gaussian, clayton, gumbel, and frank, consist of our candidate copula families. Indeed, all possible combinations of these four copulas have the ability to capture most of the possible dependence structures. We see from the Coefficient of Upper Tail Dependence (CUTD) that the mixtures that capture dependence in the upper tail are clayton gumble, gaussian gumble, and gumble frank. In accordance with the IFME, the copula mixtures are fitted with a mixture of gaussian distributions and the fitted copula parameters are obtained through a maximum likelihood estimate and displayed in Table 5.

Now we run the monte Carlo simulation with 50 000 simulations, which involves simulating a large number L of independent paths, $S_1^l, S_2^l, ..., S_n^l$ from the simulated return and estimating the option price using the average simulated discounted option payoff across the paths.

In this, we provide the simulation results for the prices of all the alternatives mentioned above based on the basket (BNP, LVMUY). On the basis of the univariate risk-neutral distributions and the calibrated copulas, the theoretical values of their prices are calculated at different values of K. In each case, we normalize the underlying to have their prices at $t = 0$ and assume a constant risk-free rate of $r = 0.03$.

Conclusion

This study proposes an approach to price multivariate options taking into account the extreme values and the dependence structure of the assets returns using mixtures of copulas and gaussian marginals. An application is made on a basket of the financial market (BNP.PA and LVMUY) downloaded from the Yahoo finance website (between January 1, 2020, and December 31, 2020).

At first, each asset's log returns are modeled by a mixture of gaussian distributions (Table 3) In all the cases, there exists a gaussian distribution with a negative mean and another one with a positive mean. one explanation is that there exist two regimes: The first one corresponds to potentially significant losses (for example, due to a financial crash), and the second one to the standard evolution of prices. The dependence between the returns is estimated using copulas mixtures.

The choice of the best fitted copulas mixture is determined by the statistic parameter of the MLE which showed that clayton gumble and gumble frank (Table 5) are the most effective at simulating dependency structures of the log returns. Prices of three options are obtained via simulation and the Clayton gumble and Clayton gaussian have relatively large values for all the options meaning that they capture more information than the other structures (Tables 6-8 and Figs. 5-6). This can also be observed from the contour plots (Figs. 7-8) which shows the strength of the dependency. Also, the differences in prices for all the copulas are very small showing that all the mixture captures the dependence structure of the returns.

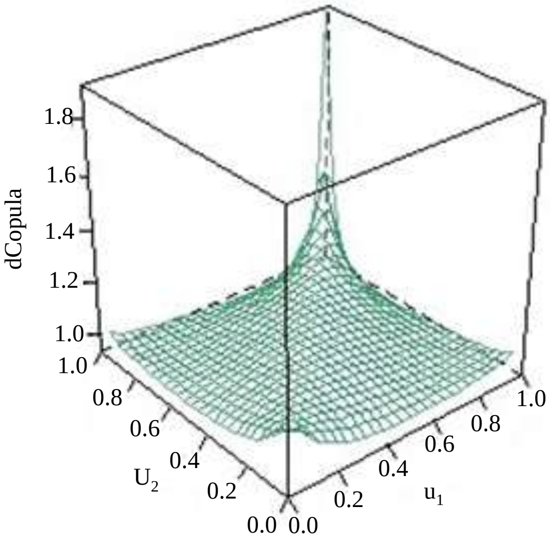

Fig. 5: Clayton gumble wireframe plot

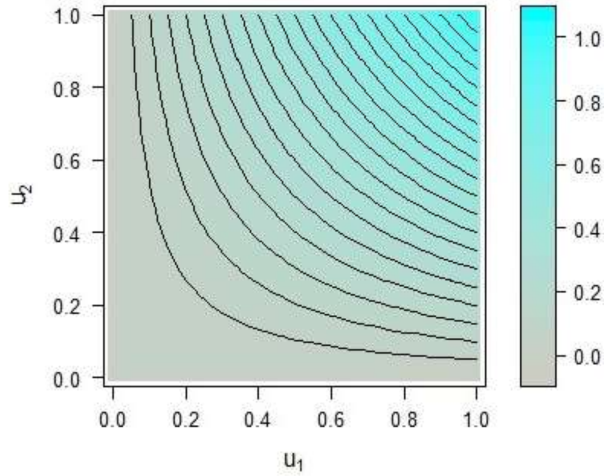

Fig. 7: Clayton gumble contour plot

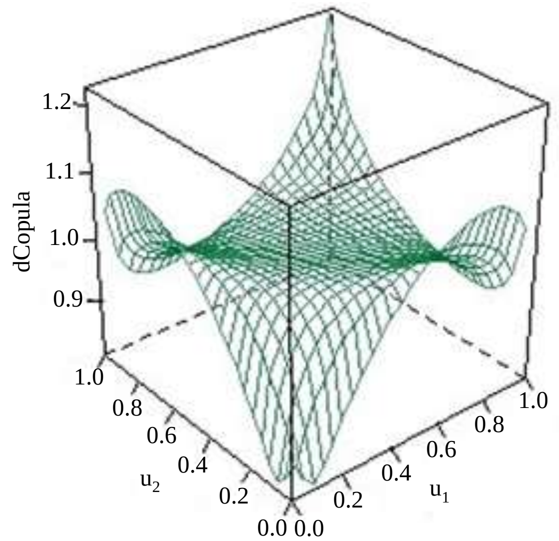

Fig. 6: Clayton gaussian wireframe plot

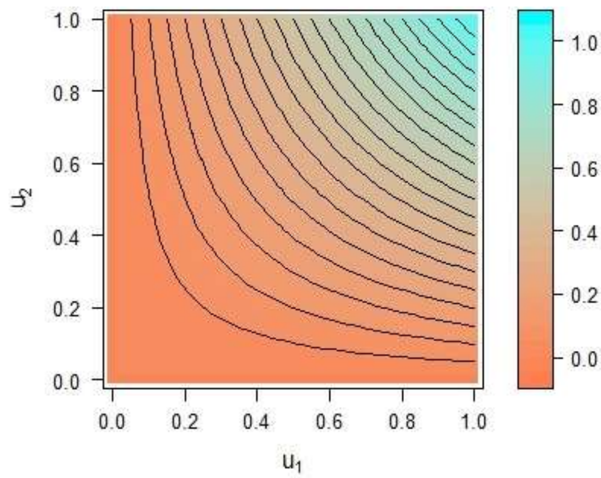

Fig. 8: Clayton gaussian contour plot

Table 6: Worst of options prices with different strikes (K)

	Clayton gumble	Gaussian gumble	Gaussian clayton	Gaussian frank	Clayton franck	Gumble franck
K = 0.5	0.3030	0.3010	0.298	0.3010	0.2950	0.313
K = 1.0	0.0710	0.0690	0.073	0.0694	0.0660	0.077
K = 1.5	0.0130	0.0126	0.014	0.0121	0.0111	0.016

Table 7: Prices of barrier options with different strikes K

	Clayton gumble	Gaussian gumble	Gaussian clayton	Gaussian frank	Clayton franck	Gumble franck
K = 1	1.283	1.2730	1.284	1.277	1.258	1.286
K = 2	0.476	0.4560	0.485	0.478	0.471	0.482
K = 3	0.127	0.1200	0.124	0.120	0.121	0.155

Table 8: Prices of the spread options with different strikes *K*

	Clayton gumble	Gaussian gumble	Gaussian clayton	Gaussian frank	Clayton franck	Gumble franck
K = 0.5	0.045	0.0385	0.0395	0.0445	0.0461	0.0371
K = 1.0	0.015	0.0170	0.0167	0.0150	0.0180	0.0160

Further research could consider the use of nonparametric or hierarchical archimedean copulas with gaussian mixture marginals. Also, the use of cross validation or numerical algebraic geometry may be used to find the optimal number of gaussian mixture components.

Data Availability

Data are available on request.

Acknowledgment

The associate editor and the anonymous reviewers are sincerely appreciated for their helpful comments and recommendations that helped us to improve the quality of this study.

Funding Information

The authors have not received any financial support or funding to report.

Author's Contributions

Jimbo Henri Claver: Drafted the research project and gave final approval of the version to be submitted and any revised version. Contributed to reviewed the critical aspect of the work and its intellectual significance.

Tatanfack Emerson: Contributed to the conception and designed of the work, the collection, analysis, and interpretation of data.

Shu Felix Che: Contributed to the written of the manuscript.

Ethics

This manuscript is the sole creation of the authors and it has not been previously published. There are no potential ethical problems because the authors have already read and approved the paper.

References

Barban, A., & Di Persio, L. (2014). Multivariate option pricing with pair-copulas. *Journal of Probability, 2014*. https://downloads.hindawi.com/archive/2014/839204.pdf

Ball, C, & Roma, A. (1994). Stochastic volatility option pricing. *Journal of Financial and Quantitative Analysis, 29(4), 589-607.* https://doi.org/10.2307/2331111

Benth, F. E., Nunno, G. D., & Schroers, D. (2022). A topological proof of Sklar's theorem in arbitrary dimensions. *Dependence Modeling*, *10*(1), 22-28. https://doi.org/10.1515/demo-2022-0103

Berton, E., & Mercuri, L. (2021). Spread Option Pricing in a Copula Affine GARCH (p, q) Model. *arXiv preprint arXiv:2112.11968*. https://doi.org/10.48550/arXiv.2112.11968

Cai, Z., & Wang, X. (2014). Selection of mixed copula model via penalized likelihood. *Journal of the American Statistical Association*, *109*(506), 788-801. https://doi.org/10.1080/01621459.2013.873366

Carley, H., & Taylor, M. D. (2002). A new proof of Sklar's theorem. *Distributions with Given Marginals and Statistical Modelling*, 29-34. https://doi.org/10.1007/978-94-017-0061-0_4

Cha, S. Y. (2010). Pricing Derivatives with Optimized gaussian Mixture. *Available at SSRN 2471477*. http://doi.org/10.2139/ssrn.2471477

Chiou, S. C., & Tsay, R. S. (2008). A copula-based approach to option pricing and risk assessment. *Journal of Data Science*, *6*(3), 273-301. https://doi.org/10.6339/JDS.2008.06(3).503

Church, C. (2011). Pricing of path-dependent basket options using a copula approach. https://www.semanticscholar.org/paper/Pricing-of-path-dependent-basket-options-using-a-Church/e396ed6b4b645c85badeca26a02853a62158a67b#paper-header

Durante, F., Fernández-Sánchez, J., & Sempi, C. (2012). Sklar's theorem obtained via regularization techniques. *Nonlinear Analysis: Theory, Methods & Applications*, *75*(2), 769-774. https://doi.org/10.1016/j.na.2011.09.006

Ghojogh, B., Ghojogh, A., Crowley, M., & Karray, F. (2019). Fitting a mixture distribution to data: Tutorial. *arXiv preprint arXiv:1901.06708*. https://doi.org/10.48550/arXiv.1901.06708

Hassane, A. M., Diakarya, B., WendKouni, Y., & Bisso, S. (2021). Pricing Multivariate European Equity Option Using Gaussians Mixture Distributions and EVT-Based Copulas. *International Journal of Mathematics and Mathematical Sciences*, *2021*, 1-9. https://doi.org/10.1155/2021/7648093

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, *6*(2),327-343 https://doi.org/10.1093/rfs/6.2.327

Hofert, M., Kojadinovic, I., Mächler, M., & Yan, J. (2018). *Elements of copula modeling with R*. Springer International Publishing. https://link.springer.com/book/10.1007/978-3-319-89635-9

DOI: 10.3844/jmssp.2023.1.12

Kim, M. S., & Kim, S. (2015). Digital Option Pricing based on Copulas with Stochastic Simulation. *The Pure and Applied Mathematics*, *22*(3), 299-313. https://doi.org/10.7468/JKSMEB.2015.22.3.299

Kouritzin, M. A., & Mackay, A. (2020). Branching particle pricers with Heston examples. *International Journal of Theoretical and Applied Finance*, *23*(01), 2050003. https://doi.org/10.1142/S021902492050003X

Malgrat, M. (2013). Pricing of a "worst of" option using a Copula method. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A668285&dswid=-2357

Moore, D. S., & Spruill, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *The Annals of Statistics*, 599-616. https://www.jstor.org/stable/2958431

Pfaff, B. (2016). *Financial risk modelling and portfolio optimization with R*. John Wiley & Sons. https://doi.org/10.1002/9781118477144

Schweizer, B., & Sklar, A. (1974). Operations on distribution functions not derivable from operations on random variables. *Studia Mathematica*, *52*(1), 43-52. https://doi.org/10.4064/SM-52-1-43-52

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, *8*, 229-231. https://doi.org/10.5023/jappstat.32.77