COPULA FUNCTIONS: A SEMI-PARAMETRIC APPROACH TO THE PRICING OF BASKET CREDIT DERIVATIVES

Marc Rousseau [1]

A THESIS SUBMITTED FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS

NATIONAL UNIVERSITY OF SINGAPORE

August 2007

[1]Ecole Centrale Paris - France - National University of Singapore; marc.rousseau@centraliens.net

1

Abstract

Le but de cette thése est de présenter la théorie des fonctions copules. Le principal intérêt de celles-ci est qu'elles permettent l'étude de la dependance entre des variables stochastiques, et plus particulièrement dans le domaine de la finance, celles-ci permettent le pricing de paniers de dérivés de crédit. Ainsi, nous commencerons par introduire les concepts fondamentaux relatifs aux fonctions copules. Ensuite, nous montrerons qu'elles sont un instrument très puissant permettant la modélisation fine de la structure de dépendance d'un échantillon de variables aléatoires. En effet, la famille des fonctions copules est trés diversifiée et chacune d'entre elles permet de décrire un certain type de structure de dépendance. Par conséquent une fonction copule peut être choisie pour décrire précisément des données empiriques. La deuxième étape de notre étude consistera à pricer un panier de dérivés de crédit. Pour ce faire, nous mettrons en place une simulation de Monte-Carlo sur un panier de CDS. La structure de corrélation des temps de défaut sera modélisée par différents types de fonctions copules.

The aim of this thesis is to present the copula function theory. Copula functions are useful to analyze the dependence between financial stochastic variables, and in particular, these methods allow the pricing of basket credit derivatives. We will first introduce the basic mathematical concepts related to copula functions. Then, we will show that they are very powerful tools in order to model the dependence structure of a random sample. Indeed, the copula function family is a very large family and each copula function depicts a certain kind of dependence structure. As a consequence, a copula function can be chosen to accurately fit empirical data.

The second step of our study will be the pricing of credit derivatives. To do so, we will perform a Monte-Carlo simulation on a basket CDS. The default correlation structure will be represented by different copula models.

Contents

List of Figures

Acknowledgments

First of all, I would like to thank Pr Oliver Chen who supervised the writing of this thesis which was a new kind of exercise for me. His patience and commitment enables me to finish this thesis despite the big distance between our two countries. I also would like to thank Pr Ephraim Clark, from the Middlesex University, who helped me in the writing of my thesis. Finally, I also thank the National University of Singapore, which permits me, through a double degree diploma with my faculty in France, to study in Singapore.

Introduction

The credit derivatives area is one of the fastest growing sectors in the derivative markets. During the first half of 2002, the notional amount of transactions has been US$ 1.5 trillions reaching US$ 8.3 trillions during the second half of 2004, compared to, respectively, US$ 2.2 trillions and US$ 4.1 trillions for the equity derivatives market. Nowadays, tranches of CDO (Collateralized Debt Obligation), for instance, are considered by traders as vanilla products.

In this thesis, we will study how copula functions can be used in mathematical finance in order to improve the accuracy of financial models. More practically we will study how copula functions prove to be very powerful tools to model the default correlation and then price financial products such as nth-to-default Basket CDS (Credit Default Swap). This credit derivative generally references 5 to 20 credits, and protects the buyer against the default of n credits, by receiving a cash amount if n credits or more default. Studying Basket CDS is a very challenging exercise because it involves correlation pricing, which is generally not easy to model. The correlation problems are inferred by the fact that all the companies are linked to each other by certain factors which are, for instance, the interest rates, the price of commodities, the political and economic situation of a country, etc. The Asian crisis in 1997 or the Internet bust in 2001 are good examples of correlation.

For approximately ten years, copula functions have become a very hot topic in the field of credit derivatives, and numerous articles have been published on that issue. However, as this field is still new compared to equity derivatives, the literature lacks basic books to learn copula and their applications from A to Z, and this literature

is made of many, very interesting articles, yet sometimes not easy to understand because they only deal with parts of the copula function theory. In this thesis, we will try to collect information through those articles and explain the main topics related to the theory of copula functions.

Before going further into the history of the discovery of the copula functions, we should first have a quick look at the reasons why they are so popular as a financial modeling tool. One of the most interesting advantages of copula functions is that this kind of function is a representation of joint distributions. As a consequence, the marginal behavior described by the marginal distributions is disconnected from the dependence, captured by the copula. Thanks to this splitting of the marginal behavior and the dependence structure, copula functions enable financial modeling where the joint normality assumption is abandoned and where more general joint distributions are used.

Historically, Sklar is one of the pioneers of copulas. In 1959, Sklar [23] introduced the concept of copulas and in its article [24] published in 1973, he proved elementary results that relate copulas to distribution functions and random variables. In particular, he considered a copula function C, and $(F_1, \cdots, F_n)$, a set of marginals, and proved the existence of a probability space where he could define a copula function C associated with a set of random variables $X_1, \cdots, X_n$ defined over that probability space. Another very important early contribution to the theory of copula functions has been provided by Frank in 1979. In his article [9], Frank's copula appeared first and were described as a solution to a functional equation problem. That problem involved finding all continuous functions such that $F(x, y)$ and $x + y - F(x, y)$ are associative. Then, in the beginning of the 90's, the canadian statistician Christian

Genest [11], [12] worked on the issue of Archimedean copula functions which will be described later in this thesis. He particularly describes methods to estimate the function which determines the Archimedean copula. Finally, in his book [21] published in 1999, Nelsen described the entire knowledge about the theory of copula functions.

As a consequence of all those fundamental researches, the field of copula functions was well defined in the second part of the 90's and some specific applications of copula functions to finance appeared. Since that date, hundreds of articles have been published applying copula functions to financial problems and more particularly to problems related to the pricing of credit derivatives. In this thesis we will pay particular attention to Li's article, [18] which describes how to use Gaussian copula functions, and Joe and Xu [16] for an estimation method of inference functions for marginals. For instance, applications of copula functions are described in Cadoux and Loizeau [4] and Gatfaoui [10].

The aim of this thesis is to present as clearly as possible a very powerful mathematical tool and present some of its applications in the financial domain. As a consequence, we will build this thesis around two aspects: on the one hand, the theory of copula function which has been described through many articles which will be studied and compiled. On the other hand, we will apply this theory to price Basket CDS. As the reader can see in the title, we will focus on the semi-parametric estimation of copula functions, which means that we will not try to estimate the parameter of a copula through, for example, a maximum likelihood estimation. However, we will use measures of concordance to determine the parameters of the copula and then try to choose the best one. We will explain all those terms and ideas throughout this thesis. Thus, in a first chapter, we will present the pricing of CDS using the hazard

rate function, which is a modeling of the default time repartition. In the second part of this first chapter, we will have a general discussion about what is correlation and how it can be estimated. This second part aims to give to the reader some basic knowledges of what is correlation, and why it is essential to study it when pricing credit derivatives. Then, chapters 2 and 3 focus on the mathematical aspects behind the copula function theory. In the beginning of chapter 2, we will define what is a copula function and see its main properties with some examples. We will also see in a second part what is correlation measurement, which will be used later in chapter 4 to perform semi-parametric estimations of the copula functions. In chapter 3, we will focus on the theory of Archimedean copula functions, which is a very widely used family of copula functions, mainly because it has very interesting properties which will be described. Finally, in the fourth chapter, we will use the results demonstrated in chapters 2 and 3 to realize two complementary applications of copula functions: the pricing of a simple Basket CDS, and the development of an algorithm which will enable us to choose the best copula regarding a dependence structure given by market data.

Before developing the introduction on the credit derivatives market, we should first go back to the title of this thesis and explain it: "'copula functions: a semi-parametric approach to the pricing of credit derivatives"'. As we will see in the following, the Archimedean copula functions we will use are parametric copula functions. However, we will use a terminology close to the terminology presented by Genest and Rivest [12] which consider that the estimation method is semi-parametric compared for example to the Maximum Likelihood Method which aims to estimate the parameter of the copula function by maximising the likelihood depending on

the parameters of the copula. Indeed, in our study, we will estimate the empirical copula of our dataset which is a non-parametric copula. However, in order to be able to model our dependence structure, we will then describe a method to find the copula function which will describe the dependence structure the most accurately, but we will never directly estimate a parametric copula function. In order to better understand this point, we will develop it in 3.3.1.

On the credit derivatives market

Before focusing on the issue of Basket-CDS (Credit Default Swap), we will first try to have a broader view on the issue of the credit derivatives market, which, as we saw before, is a very fast growing market. Mainly, the goal of this market is to transfer the risk and the yield of an asset to another counterpart without selling the underlying asset. Even if this primary goal might has been turned away by speculators, banks remain the main actor of this market in order to hedge their credit risk and optimize their balance sheet.

In order to understand why credit derivatives are very useful to banks, we can look at a simple example. Consider two banks Wine-bank and Beer-Bank. Wine-bank is specialized in lending money to wine-producers whereas Beer-Bank is specialized in lending money to brewers. As a consequence, both banks have a portfolio of one type of credits, either correlated to the health of the wine sector or the beer sector. The other consequence of this specialization is that both banks have been able to develop a very good knowledge of its sector, thus they are able to lend money at a better rate, because they are able to determine the credit risk much more accurately than if both had to look at both sectors without being able to develop thorough

knowledge of its sector. To summarize, we can say that both banks are able to select the best companies in each sector, compared to the situation where each bank would lend money to either wine-producers or brewers. However, the main problem concerning this segmentation of the market is that if tomorrow, consumption of wine decreases sharply in favor of the consumption of beer, Wine-bank could face more credit defaults even if its portfolio is only made of good vineyards (financially speaking!). As both banks have the same risk of facing an increase of defaults in its sector because of external causes, they will try to hedge that risk. Intuitively, we can understand that the main problem of both banks is that they have not diversified their portfolios. One of the possibilities would be for both to sell part of their portfolio to each other. As a consequence, both banks would be hedged against the decline of one beverage as far as the lost of consumption of one beverage is supposed to be offset by an increase of consumption of the other beverage. However, the main problem of this method is that a client probably won't be very happy to know that even if he has signed a contract with bank A, his contract has been sold to another bank. Besides, this transaction implies the exchange of the notional of each contract, and as a consequence, the sale will not be easy to achieve. That's why researchers have imagined another way to transfer the risk linked to a credit, without transferring the credit itself. This category of products is named credit derivatives, in opposition to the products derived from the bond family, which are the underlying assets of interest rates derivatives (likewise, stocks are the underlying assets of equity derivatives).

We have already mentioned credit default swaps (CDS) before. Indeed, this product is becoming more and more popular and its aim is to hedge the potential loss related to a credit event. More precisely, the CDS is a contract signed between

two counterparts. The buyer of the CDS agrees to pay regularly a predetermined amount to the seller of the CDS. On the other hand, in case of a credit event (like a default, for example, but the notion of credit event can be broader, depending on the contract), the seller agrees to reimburse the buyer of any losses caused by this credit event. Nowadays, a similar product of CDS has developed, the Collateralized Debt Obligation, which can be a structured like a basket CDS.

1 Preliminary Results and Discussions

Before studying precisely the theory of copula functions, we first introduce some basic results which will be used in the coming chapters. After the presentation of the hazard rate function which is a very simple tool representing the instantaneous default probability for an asset which has survived until the present time, we will present how it can be used to price a CDS. Then, we will have a short discussion on what default correlation is and how it can be measured.

1.1 The Hazard Rate Function

In this subsection, we want to model the probability distribution of time until default. We denote T the time until default and thus study the distribution function of T. From this distribution function, we derive the hazard rate function. These hazard rate functions will be used to calculate the price of the 1st-to-default Basket CDS.

Let $t \to F(t)$ be the distribution function of T

$$F(t) = \mathbb{P}\left[T \leq t\right], \qquad t \geq 0. \tag{1}$$

Let $t \to S(t)$ be defined by

$$S(t) = 1 - F(t) = \mathbb{P}\left[T > t\right], \qquad t \geq 0. \tag{2}$$

The function $t \to S(t)$ is called the survival function, and it gives the probability that a security will attain age t.

We assume that $t \geq 0$ and $S(0) = 1$. Let $t \to f(t)$ be the probability density function of $t \to F(t)$

$$f(t) = F'(t) = -S'(t) = lim_{\Delta \to 0^+} \frac{\mathbb{P}\left[t \leq T < t + \Delta\right]}{\Delta}. \tag{3}$$

At this step, we have defined the distribution function and the probability density function of the survival function of our asset. We will now introduce the hazard rate function, which gives the instantaneous default probability for an asset which has survived until time x.

Consider the definition of the conditional probability. Assume A and B are two events

$$\mathbb{P}\left[A|B\right] = \frac{\mathbb{P}\left[A \cap B\right]}{\mathbb{P}\left[B\right]}.$$

Thus

$$\begin{aligned} \mathbb{P}\left[x < T \leq x + \Delta x | T > x\right] &= \frac{\mathbb{P}\left[x < T \leq x + \Delta x\right]}{\mathbb{P}\left[x < T\right]} \\ &= \frac{S(x) - S(x + \Delta x)}{S(x)} \\ &\approx \frac{f(x)\Delta x}{S(x)}. \end{aligned}$$

Finally, we define h, the hazard rate function[2] as

$$h(x) = \frac{f(x)}{S(x)} = -\frac{S'(x)}{S(x)}. \tag{4}$$

The hazard rate function is the probability density function of T (the time at which the default occurs), at the exact age x, given survival until x.

---

[2]This function is also called the default intensity

In (4), we can recognize a first order ordinary differential equation, so that

$$S(t) = e^{-\int_0^t h(s)ds}.$$

And

$$f(t) = h(t)e^{-\int_0^t h(s)ds}, \qquad t \geq 0. \tag{5}$$

If we make the assumption that $\forall t$, $h(t) = h$, with h constant

$$f(t) = he^{-ht}, \qquad t \geq 0.$$

We can recognize the probability density function of an exponential distribution $F(t) = 1 - e^{-ht}$, with $\mathbb{E}(T) = \frac{1}{h}$ [3] and $V(T) = \frac{1}{h^2}$. The skewness of this distribution is equal to 2 and its kurtosis is equal to 9.

Finally, we want to determine the price of the 1st-to-default Basket CDS. The method that is described will be used later to derive the price of a first to default basket CDS from Monte-Carlo simulation. Before going further, we need to understand that this method is only valid if the default time repartition can be modeled by an explicit hazard rate function. To illustrate this method, we assume that the hazard rate function is constant: $\forall x$, $h(x) = h$. Let V be the value of our 1st-to-default Basket CDS, P the payoff of the basket CDS, and T_d the time until maturity of the basket CDS. Let $R \in [0,1]$ be the recovery rate, ie the amount of money which

---
3

$$\mathbb{E}(T) = h \int_0^\infty Te^{-hT}dT = \frac{1}{h}.$$

will proceed from the reimbursement of the credit after the default event, and r the interest rate, which is assumed to be constant. Then

$$
\begin{aligned}
V &= (1 - R) \int_0^{T_d} P e^{-rt} f(t) dt \\
&= (1 - R) h \int_0^{T_d} P e^{-[(r+h)t] dt} \\
&= (1 - R) \frac{h}{r + h} P (1 - e^{-(r+h)T_d}).
\end{aligned}
\tag{6}
$$

This formula is valid if h is constant. However, we can also consider hazard rate functions which are piecewise-constant, so that, if $h(t) = \sum_{k=0}^{N} h_k(t)$, with $h_k(t) = h_k$ if $t \in [k, k+1]$ and 0 otherwise. If we denote the time until maturity of the first default T_d to be equal to $N + 1$, then, the price of our 1st-to-default Basket CDS is determined by

$$
V = (1 - R) \sum_{k=0}^{N} \frac{h_k}{r + h_k} P e^{-(r+h_k)T_d}.
\tag{7}
$$

This approach is a theoretic approach of the pricing of a 1st-to-default Basket CDS, as we generally do not know the closed formula of h, the hazard rate function. As a consequence, it can not be used directly to price nth-to-default Basket CDS. However, this result will be used in part 4.1.3 in order to derive the price of a basket CDS using a Monte-Carlo simulation.

1.2 The pricing of CDS

The study of credit derivatives is a very broad issue. To compare it with equity derivatives, we can see CDS as similar to call and put options, in the way that both are basic instruments used to build more sophisticated strategies based on derivatives.

Indeed, CDS is the most basic credit derivative and is generally the first component of a more complicated credit derivative like synthetic CDO. Thus, in this section, we will see a first analytical method that allows to derive the price of this CDS. In our study, we will use the seller convention, so that we will study the case of the seller of the protection. Thus, we will be able to define our profit expectation which will be called Feeleg, and our loss expectation which will be called the Defleg.

As we stated before, the CDS is described by its a maturity which is the time until maturity of the contract. During this period, several events will occur. Each month for instance, the buyer of the protection will pay the seller a fixed amount which will be called the spread of the CDS. This amount is generally determined as a percentage (in basis points) of the notional amount of the CDS. In order to simplify our study we assume that:

- The payments related to the CDS and made by the buyer of the protection are made at discrete times (every months for instance) at T_i.

- If we denote by R the recovery rate and C_{CDS} the spread (or cost) of the CDS, the money exchanged at each time t is equal to C_{CDS} for the seller of the protection if no default has occurred before time t (with the probability $1 - F(t) = e^{-\int_0^t h(s)ds}$), and $1 - R$ for the buyer if there is a default at time t (with a probability $h(t)e^{-\int_0^t h(s)ds}$) .

- p is the number of payments of the spread.

We now derive the price of a CDS using these hypothesis and the definition of the hazard rate function presented in the section before.

Let N denote the nominal amount of the portfolio, r the riskless interest rate, τ the time of a credit default and C_{CDS} the spread used for the pricing of the CDS.

$$Feeleg(CDS) = \mathbb{E}_Q \left[C_{CDS} * N * \sum_{i=1}^{p} (T_i - T_{i-1}) * 1_{\tau > T_i} * e^{-\int_0^{T_i} r(t)dt} \right].$$

With $\mathbb{E}_Q \left[1_{\tau > T_i} \right] = S(T_i)$ the survival probability until T_i, and $t \to r(t)$ the risk-free interest rate function at time t.

And $S(T_i) = e^{-\int_0^{T_i} h(t)dt}$, with $h(t)$ the instantaneous default intensity at date t. Here, we will suppose that h is a continuous and deterministic function of time.

Thus, with C_{CDS} the price (or spread) of the CDS, we have:

$$Feeleg(CDS) = C_{CDS} * N * \sum_{i=1}^{p} (T_i - T_{i-1}) * e^{-\int_0^{T_i} r(t)dt} * e^{-\int_0^{T_i} h(t)dt}.$$

Similarly, we have:

$$Defleg(CDS) = N * (1 - R) * \sum_{i=1}^{p} \left[(S(T_{i-1}) - \mathbb{P}(0, T_i)) * e^{-\int_0^{T_i} r(t)dt} \right].$$

Finally, we can calculate the Net Present Value, or NPV, as the difference between the profit expectation and the loss expectation:

$$NPV(CDS) = Feeleg(CDS) - Defleg(CDS).$$

We can now define the fair spread or implied spread of the CDS as the spread which will set the value of the contract to 0 at the time of the transaction, using R the recovery rate, we obtain:

$$NPV(CDS) = 0,$$

$$C_{CDS} = \frac{N * (1 - R) * \sum_{i=1}^{p} \left[(S(T_{i-1}) - S(T_i)) * e^{-\int_0^{T_i} r(t)dt} \right]}{N * \sum_{i=1}^{p} \left[(T_i - T_{i-1}) * S(T_i) * e^{-\int_0^{T_i} r(t)dt} \right]}.$$

To conclude, we have derive in this sub-section the price of a CDS, using the concept of the hazard rate function we introduced previously. However, it is very important to understand that the main problem in pricing CDS is not a default correlation problem but a default time modeling problem. Even if the default time modeling is not the core problem developed in this thesis, it is necessary to understand where the frontier lies. In the following, we will mainly focus on correlation modeling problem which arises when we mix within a portfolio several CDS together.

1.3 On Default Correlation

The focus on the correlation problem is not something new in finance. Indeed, correlation is widely studied in order to understand the behavior of portfolios and indices in particular, and more generally to understand any problem where the payoff depends on more than one parameter or instrument.

The first question one should ask when confronted with correlation is: *what is correlation?* According to JP Morgan, it is the *"'strength of a relationship between two or more variables"'*[19]. The most well-known correlation is the Pearson correlation. However, several other kinds of non-linear correlation exist. Besides the polynomial or log correlations, other techniques such as Spearmann or Kendall rank correlation coefficients are also used as they provide a method which can overcome some of the problems which can be encountered when using linear correlation calculations. However, this rank correlation coefficient method is not widely used compared to the most common method of calculating correlation which is based on the Pearson coefficient defined by:

$$\rho = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}.$$

With X_i and Y_i the observations, and $\bar{X}$ and $\bar{Y}$ the means of the random variables X and Y.

Intuitively, we understand this correlation coefficient when it is equal to $1, -1$ or 0, which respectively mean that if the correlation coefficient is equal to 1, then the data are perfectly correlated, if it is equal to -1, then the data are perfectly

negatively correlated, and finally, if the correlation coefficient is equal to 0, then the data are independent. However, the main problem in the interpretation of this coefficient arises when it is not equal to one of this three figures. How can we interpret a correlation coefficient of 0.6, or -0.3? Obtaining such figures, we cannot actually state if the data are indeed correlated or not. One can suggest that an $80 - 20$ rule can be applied, which states that a correlation coefficient beyond 80% means that the data are highly correlated, whereas a correlation coefficient under 20% means that the data show little or no correlation.

However, the first thing one should examine carefully before performing a correlation calculus is the relevance of such a calculus. Indeed, looking at the correlation between the profits generated by a french car maker and a retail bank in Singapore will give a result, mathematically speaking, but is it really relevant for drawing a conclusion? Probably not. Thus, one of the first things we will have to examine carefully is probably not which correlation method must be used to calculate a correlation, but rather if the calculus has any consistence.

As we can see everyday, correlation is all around us: we can study the correlation between the size of men and their birth dates, the revenue of a family and the number of cars they own, and the profits generated by a bank in France and in Singapore. A full study of the theory of correlation is not the subject of this thesis, and that is why we will concentrate our study on the subject of default correlation.

1.4 Estimating default correlation

In the preceding sections, we have seen that default correlation is a key point in pricing any portfolio of credit derivatives. So that we now study three methods to estimate this correlation.

1.4.1 Estimating default correlation from historical data

Historical estimation of the default correlation between two companies is not something easy to realize, and if we want to look at it, it is probably because these very companies still exist and have not defaulted before. Unlike historical volatility, for instance, historical default correlation is not something easy to observe.

For stand alone companies, it is relatively easy to identify the rate of default within an entire sector or even the entire market. However, it is not easy to draw any conclusions from those data. For example, the fact that lots of businesses are dependent on the business cycle makes the job even harder because if we don't look at a very long period, and draw conclusions on another very long period, it is very easy to make false conclusions.

However, one very useful method using historical data is derived from the historical default data provided by rating agencies such as Standard and Poors or Moodys, which gives the probability of default during a period as a function of the rating of a company. These probabilities of default are in fact historical probabilities of default as they are based on the observations made by the rating agency.

1.4.2   Estimating default correlation from equity returns

Compared to the historical estimation of the default correlation, estimating correlation from equity returns is commonly used as a method to price basket credit derivatives. For example, the CreditMetrics model is based on the Merton framework which suggests that credit and equity are related. Indeed, considering the position of an equity and a bond holder in terms of options, an equity holder can be considered as long a call option on the assets of the firm whereas the bond holder can be considered as short a put on the same assets. As a consequence, using the put-call parity, we can conclude that equity and debt are related. Moreover, if you consider that the assets of a company are represented by a random variable, then the company will default at some threshold (which can be for example when the assets are worth strictly less than debt). However, as a time-series of the assets of a firm is not something very easy to get, CreditMetrics uses equity returns as a proxy. Thus, we can determine the correlation between two-firms as the correlation between their equity returns and using the Merton [20] assumption we can finally say that in order to estimate the default correlation between two firms, the correlation between the equity of those two firms can be used as a proxy.

The main advantage of this method is that it is relatively easy to implement as the stock prices are something very easy to get. This approach will be used in part 4.4 in order to approximate the default correlation of a portfolio based on equity returns correlation.

1.4.3 Estimating default correlation from credit spreads

In the previous section, we used the stock price for which historical data are generally easy to get. In this section, we will use other market data which are corporate bond prices. Indeed, we know that bond prices include two components: the interest rate and the credit risk related to a particular company. Since interest rates are observable, we can strip them out of the price of bonds, then only the credit component of the bond remains. From this component it is then easy to estimate the default probability as soon as we can estimate the recovery rate (which is given by the amount of money a bond holder expects to get in case of a default). This estimation of the recovery is the tricky part because it is not easy to estimate how much will be redeemed by bond holders in the case of default. However, we can use data from rating agencies which give an estimate of the recovery rate based on historical observations. Another problem raised by the estimation of default probability from bond prices is that depending on the liquidity of the bond and depending on other technical reasons inferred by the market quotation, bond prices can be polluted by a third component which would be a market component and which is not easy to eliminate.

Another way of determining the default probability is to use the credit default swap (CDS) market which has become an efficient market with several years of history. Indeed, with CDS, it is possible to directly convert the spread quoted in the market into a default probability. However, the technical problems due to the quotation are not eliminated.

1.5 How to trade correlation?

Correlation trading is based on new financial products such as DJ Tranched TRAC-X which is a standard CDO. Indeed, pure correlation traders buy or sell a tranche of a synthetic corporate CDO, with the view that the correlation over the period they will hold the CDO will be different to that embedded in the instrument.

The world of credit derivatives is evolving very quickly: instruments that were considered exotic a few years ago are now regarded as vanilla products. In 2002 a tradable synthetic index was created (iBoxx Diversified index in Europe and Dow Jones CDX index in North America), a product that brought high volume, low-margin trading to credit for the first time and provided a new way to take or hedge exposure to the broad credit market.

Nowadays, the most traded correlation related products are: Nth-to-default basket (Standard or Tailor-made); Single-tranche synthetic CDOs; CDO-squared (CDO of CDO); Index tranches.

The most basic of the correlation-based products are those related to baskets of credits, with the first-to-default (FTD) basket the most familiar of these. Investors in FTD structures sell protection on a reference portfolio of names and assume exposure to the first default to take place within the pre-defined basket of credits. On occurrence of a default, the FTD swap works like an ordinary CDS. First-to-default baskets typically offer credit exposure to between three and 10 companies. From the perspective of investors, the principal interest of a FTD structure is that it offers a higher yield than any of the individual credits within the basket and limits downside risk in the event of default. An additional interest for investors is the transparency of

FTD swaps, because the credits included within the basket are generally the choice of the investor.

2 Some Insights On Copula Function

Copula functions are a very powerful but also new tool, which was discovered in the late 60's. The application of copula functions to financial problems began in the 80's. As we saw in the introduction, one of the main problems related to the pricing of credit derivatives is the implementation of the correlation between the different assets of a portfolio. Particularly in the environment of financial markets where random variables are poorly described by Gaussian distributions, the use of more precise models describing the behavior of portfolios is nowadays something of paramount importance. Thus the utilization of copula functions has become something quite common in financial mathematics in order to model the dependence structure of a portfolio. Indeed, the main interest of copula functions is that they make it possible to dissociate the dependence structure of multiple assets from their marginal distribution. As a consequence, we can imagine studying the very common situation where a portfolio marginal distribution is modeled by a Student-t distribution whereas the dependence structure is described by a Gaussian distribution. The flexibility also enables one to easily study another situation where the dependence structure (ie the multivariate distribution function) is not Gaussian, and thus take into account the fact that the dependence structure of a portfolio can have a fat tail, which means that extreme values are more likely to happen than what the Gaussian dependence structure describes. Finally, as copula functions split the problem of the estimation of the marginals and the dependence structure, they are generally more tractable and easier to estimate than multivariate distributions which are not described in terms of copula functions.

In this introduction, we will present the basic definitions and theorems describing the world of copula functions. Our aim will be to try to understand mathematically what copula functions are. This chapter will then enable us to apply the utilization of copula functions to financial problems. After studying some mathematical definitions and theorems which will be used later, we introduce the definition of correlation measurement which is a very powerful tool to estimate copula functions.

Abe Sklar discovered copula function in 1959 as he thought that the determination of the set of the copula functions C is easier than the determination of $slF(\mathbf{F}_1, \cdots, \mathbf{F}_n)$ which is the Frechet class:

Les copules sont en général d'une structure plus simple que les fonctions de répartition (Sklar [23], page 231)[4].

2.1 Definition and Properties

In this first section, we will present some definitions and properties which will be used to describe mathematically the construction of a copula function. We particularly present Sklar's Theorem which is the basis of the copula function theory. All these definitions can be found in [3] or [22].

Let X and Y be two random variables, with F and G their respective distribution functions

$$F(x) = \mathbb{P}[X \leq x],$$

---

[4]copula generally have a more simple structure than distribution functions

$$G(x) = \mathbb{P}[Y \le y].$$

And the joint distribution function

$$J(x, y) = \mathbb{P}[X \le x, Y \le y].$$

For each pair (x, y), we can associate 3 numbers: $F(x)$, $G(y)$ and $J(x, y)$. Moreover, each number is in $[0, 1]$. Thus, each pair (x, y) has an image $(F(x), G(y))$, in the unit square $[0, 1] \times [0, 1]$, and this pair corresponds to a value of $J(x, y)$ in $[0, 1]$. The connection between F, G and J, ie between the joint distribution function and its marginal distribution functions is called a copula function.

We denote $\mathbb{R}$ the space of real numbers $(-\infty, +\infty)$, $\bar{\mathbb{R}}$, the extended space of real numbers $[-\infty, +\infty]$. A rectangle B of $\bar{\mathbb{R}}^m$ is the cartesian product of m-closed intervals

$$B = [x_{11}, x_{12}] \times [x_{21}, x_{22}] \times \cdots \times [x_{m1}, x_{m2}].$$

The vertices of B are the points (x_{i2}, x_{i1}), (x_{i1}, x_{i2}), (x_{i1}, x_{i1}), (x_{i2}, y_{i2}), $\forall i \in [0, m]$. The unit square is the product $I \times I \times \cdots \times I$, with $I = [0, 1]$. A m-dimensional real function H is a function whose domain is a subset of $\bar{\mathbb{R}}^m$, and whose image is a subset of $\mathbb{R}$.

Definition

In the case of an m-dimensional copula function, we define for a given $t \in S_1 \cdots S_m$ where the S_k are m non empty sets which have at least one element:

$$V_H(B) = \Delta_{a_m}^{b_m} \Delta_{a_{m-1}}^{b_{m-1}} \cdots \Delta_{a_2}^{b_2} \Delta_{a_1}^{b_1} H(t),$$

with

$$\Delta_{a_k}^{b_k} H(t) = H(t_1, \cdots, t_{k-1}, b_k, t_{k+1}, \cdots, t_m) - H(t_1, \cdots, t_{k-1}, a_k, t_{k+1}, \cdots, t_m).$$

Definition A m-dimensional real function H is said to be "'m-increasing"' if for all rectangles B whose vertices are in $Dom(H)$,

$$V_H(B) \geq 0.$$

Definition Given a function $H : \bar{\mathbb{R}}^n \to \mathbb{R}$ and given $S_1, \cdots, S_m = Dom(H)$ where each S_k has at least one a_k, we say that H is grounded if $H(t) = 0$ for all t in $Dom(H)$ such that $t_k = a_k$ for at least one k.

We recall that a copula function is a function that links univariate marginals (obtained with credit curves for example), to the multivariate distribution. In our thesis, the problem is to study the behavior of a portfolio (ie the multivariate distribution), knowing the univariate marginals. As we will see later, the copula is the analytical representation of the dependence structure.

Definition: For m uniform random variables, $U_1, U_2, \cdots, U_m$, the copula function is defined as a function from $[0, 1]^m \to [0, 1]$ which satisfies:

1. $C(u_1, u_2, \cdots, u_m)$ is m-increasing,

33

2. C is grounded,

3. $C(1, \cdots, u_k, \cdots, 1) = u_k \qquad \forall k \in [0, m]$,

4. $C(u_1, u_2, \cdots, u_m) = \mathbb{P}\left[U_1 \leq u_1, U_2 \leq u_2, \cdots, U_m \leq u_m\right]$.

Copula functions can be used to link marginal distributions with a joint distribution. For a given set of univariate marginal distribution functions $F_1(x_1), F_2(x_2), \cdots, F_m(x_m)$, the function $F(x_1, x_2, \cdots, x_m) = C(F_1(x_1), F_2(x_2), \cdots, F_m(x_m))$ describes the joint distribution of F.

When using copula functions, the most interesting and important theorem is the Sklar theorem [23], which establishes the converse of the previous equality:

Sklar's Theorem: If $F(x_1, x_2, \cdots, x_m)$ is a joint multivariate distribution function with univariate marginal distribution functions $F_1(x_1), F_2(x_2), \cdots, F_m(x_m)$, then there exists a copula function $C(u_1, u_2, \cdots, u_m)$ such that

$$F(x_1, x_2, \cdots, x_m) = C(F_1(x_1), F_2(x_2), \cdots, F_m(x_m)). \qquad (8)$$

Moreover, if each F_i is continuous, then C is unique.

We now denote $c(\cdot)$ the density function associated with the copula function $C(\cdot)$, we obtain c by calculating:

$$c(u_1, \cdots, u_m) = \frac{\partial^m C(u_1, \cdots, u_m)}{\partial u_1 \cdots \partial u_m}. \qquad (9)$$

If we note $f(\cdot)$ the joint density associated with $F(\cdot)$, and f_k the kth marginal density, we can show that:

$$f(x_1, \cdots, x_m) = c(u_1, \cdots, u_m) \prod_{k=1}^{m} f_k(x_k). \tag{10}$$

Thus, in this decomposition, c represents the dependence structure of $f(\cdot)$

The main advantage in using copula functions is that it allows to build complex multidimensional distributions, thanks to Sklar's Theorem, which links univariate marginals to their full multivariate distribution thereby separating the dependency structure C. When dealing with complex analytical expressions of multidimensional distributions, copula functions enable us to get more tractable expressions. Moreover, copula functions allow us to use a different marginal distribution for each asset.

Another very interesting property of the copula function, is that these functions are invariant under strictly increasing, continuous transformations. Again, we use Nelsen's book for the proof of this theorem:

Let X and Y be continuous random variables with copula C_{XY}. If α and β are strictly increasing functions on the domain described respectively by X and Y, then

$$C_{\alpha(X)\beta(Y)} = C_{XY}.$$

Indeed, let F_1, G_1, F_2, G_2 denote the distribution functions of X, Y, $\alpha(X)$, $\beta(Y)$ respectively. Since α and β are strictly increasing, $F_2(x) = \mathbb{P}\left[\alpha(X) \leq x\right] = \mathbb{P}\left[X \leq \alpha^{-1}(x)\right] =$

$F_1(\alpha^{-1}(x))$, and likewise $G_2(y) = G^{-1}(\beta^{-1}(y))$. Thus for any x, y in $(Dom(\alpha), Dom(\beta))$,

$$C_{\alpha(X)\beta(Y)}(F_2(x), G_2(x)) = \mathbb{P}\left[\alpha(X) \leq x,\ \beta(Y) \leq y\right]$$
$$= \mathbb{P}\left[X \leq \alpha^{-1}(x),\ Y \leq \beta^{-1}(y))\right]$$
$$= C_{XY}(F_1(\alpha^{-1}(x)), G_1(\beta^{-1}(y)))$$
$$= C_{XY}(F_2(x), G_2(y)).$$

We have similar results if α and β are monotonic:

- If α is strictly increasing and β is strictly decreasing, then

$$C_{\alpha(X)\beta(Y)}(u, v) = u - C_{XY}(u, 1 - v);$$

- If α is strictly decreasing and β is strictly increasing, then

$$C_{\alpha(X)\beta(Y)}(u, v) = v - C_{XY}(1 - u, v);$$

- If α and β are both strictly decreasing, then

$$C_{\alpha(X)\beta(Y)}(u, v) = u + v - 1 - C_{XY}(1 - u, 1 - v).$$

The main advantages of these results in the study of nth-to-default basket CDS, and more generally for any financial problem, is that we can study either price series or log-price series with the same copula.

The last concept that will be introduced concerning copula functions is the tail dependence, which plays a fundamental role in the description of the dependence

structure and hence the copula function. A copula $C(u,v)$ is said to have a left (lower) tail dependence if

$$\lim_{u \to 0} \frac{C(u,u)}{u} = \lambda > 0.$$

The right (upper) tail dependence is defined using the survival copula $\bar{C}$ which is defined by $\bar{C}(u,v) = 1 - u - v + C(u,v)$, and the right dependence λ_r verifies

$$\lim_{u \to 1} \frac{\bar{C}(u,u)}{1-u} = \lambda_r > 0.$$

This tail dependence enables us to directly measure the probability that two extreme events happen at the same time. This concept is used in the study of the contagion of crisis between markets or countries.

2.2 Examples of Copula Function

After having studied the main properties which characterize the copula functions, we will now introduce some of the most widely used copula functions. All these examples can be found in Jouanin et al. [17]. The first copula function we study is already known as a multivariate distribution, but probably not as a copula function. Indeed, the Gaussian copula function can be studied as a Gaussian multivariate distribution. Besides, the Student copula function tends to describe a similar dependence structure when the degrees of freedom of this distribution increases. Finally, we will introduce the empirical copula function which will be used in part 4.4 in order to determine the most appropriate copula function given an empirical set of data.

2.2.1   The Multivariate Normal Copula

This copula function is the most widely used copula function, because it is a relatively tractable copula that fits well with the Monte-Carlo simulation model.

Let Σ be a symmetric, positive definite matrix with the diagonal terms equal to 1, and ϕ_Σ the multivariate normal distribution function, with the correlation matrix Σ. Then we can define the multivariate normal copula function as

$$
\begin{aligned}
&C(F_1(x_1), F_2(x_2), \cdots, F_m(x_m); \Sigma) \\
&\quad = \phi_\Sigma(\underbrace{\phi^{-1}(F_1(x_1))}_{Y_1}, \underbrace{\phi^{-1}(F_2(x_2))}_{Y_2}, \cdots, \underbrace{\phi^{-1}(F_m(x_m))}_{Y_m}),
\end{aligned}
\tag{11}
$$

with ϕ^{-1} the inverse of the cumulative probability distribution of a Normal distribution

Moreover, the density of the Gaussian copula function is given by[5]

$$
c(u_1, \cdots, u_m; \Sigma) = \frac{1}{|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}\varsigma^*(\Sigma^{-1}-\mathbb{I})\varsigma},
\tag{12}
$$

with ς the vector of coordinates $(\varsigma_n)^*$ and $\varsigma_n = \phi^{-1}(u_n)$

Finally, Embrechts et al. in [6] have demonstrated that the Gaussian copula has no tail dependence(page 18-19).

---

[5]The symbol * mean the transpose of the vector

2.2.2 The Multivariate Student-t Copula

Let Σ be a symmetric, positive definite matrix with the diagonals terms equal to 1, and $T_{\Sigma,\nu}$ the multivariate Student-t distribution function[6], with ν degrees of freedom, with the correlation matrix Σ. The multivariate Student-t copula is defined by

$$
\begin{aligned}
C(F_1(x_1),&F_2(x_2),\cdots,F_m(x_m);\Sigma;\nu)\\
&= T_{\Sigma,\nu}(t_\nu^{-1}(F_1(x_1)),t_\nu^{-1}(F_2(x_2)),\cdots,t_\nu^{-1}(F_m(x_m)))
\end{aligned}
\tag{13}
$$

with t_ν^{-1} the inverse of the univariate Student-t distribution [7]. The corresponding density is

$$
\begin{aligned}
c(F_1(x_1),&F_2(x_2),\cdots,F_m(x_m);\Sigma)\\
&= |\Sigma|^{-\frac{1}{2}}\frac{\Gamma\left(\frac{\nu+m}{2}\right)\left[\Gamma\left(\frac{\nu}{2}\right)\right]^m\left(1+\frac{1}{\nu}\zeta^T\Sigma^{-1}\zeta\right)^{-\frac{\nu+m}{2}}}{\left[\Gamma\left(\frac{\nu+1}{2}\right)\right]^m\Gamma\left(\frac{\nu}{2}\right)\prod_{n=1}^m\left(1+\frac{\zeta_n^2}{\nu}\right)^{-\frac{\nu+1}{2}}}.
\end{aligned}
\tag{14}
$$

Given a multivariate Gaussian vector $Y = (Y_1,\cdots,Y_n)$ following a multivariate normal distribution with the correlation matrix equal to Σ, the vector ΘY is said to be Student-t distributed with n degrees of freedom if $\Theta = \sqrt{\frac{n}{X}}$, with X following a $\chi^2(n)$ law, and Θ independent of Y.

---

[6]$T_{\Sigma,\nu}(x) = \dfrac{\Gamma\left(\frac{\nu+m}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)((n\Pi)^{m/2}|\Sigma|^{1/2}\left[1+\frac{1}{\nu}x^T\Sigma x\right]^{\frac{\nu+m}{2}}}$

[7]$t_\nu(x) = \dfrac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\Pi}}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$

2.2.3 The Fréchet Bounds

We say that the copula C_1 is smaller than the copula C_2, and we write $C_1 \prec C_2$ if

$$\forall (u_1, \cdots, u_m) \in I^m, \quad C_1(u_1, \cdots, u_m) \leq C_2(u_1, \cdots, u_m). \tag{15}$$

Two specific copulas play a particular role, the lower and the upper Fréchet bounds C^- and C^+

$$C^-(u_1, \cdots, u_m) = max \left(\sum_{p=1}^{m} u_p - m + 1, 0 \right); \tag{16}$$

$$C^+(u_1, \cdots, u_m) = min(u_1, \cdots, u_m). \tag{17}$$

It can be shown [8] that the following order is verified for any copula C

$$C^- \prec C \prec C^+. \tag{18}$$

The 3D graphs in Figure 1, illustrates respectively the minimum (C^-) and the maximum (C^+) copulas for the two-variable case.

2.2.4 The Empirical Copula

A copula function can also be calculated empirically. Indeed, a cumulative distribution function F of a random variable X can be written empirically from a sample $(x_1, \cdots, x_n)$ of N realizations of X by the function:

$$F_e = \frac{number\ of\ x_i\ such\ that\ x_i \leq x}{N}.$$

---

[8] See Nelsen (1998), Theorem 2.2.3

40
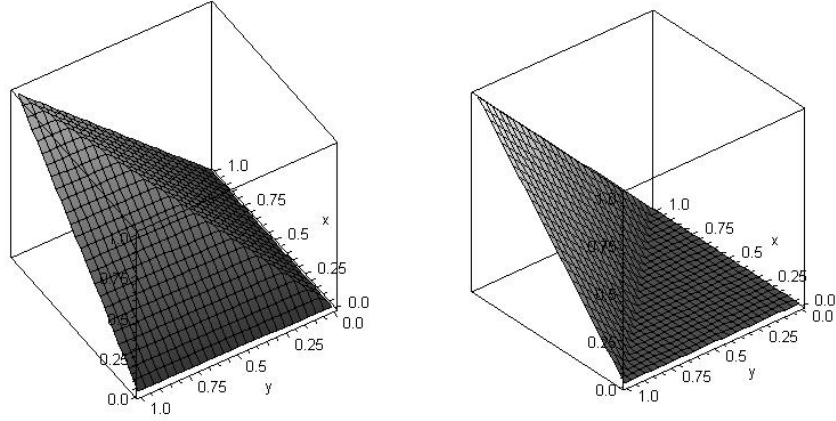
Figure 1: Representation of the minimum (left) and maximum (right) Fréchet copula

In the following, we will use the following notation:

$$F_e = \frac{\#x_i \mid x_i \leq x}{N}.$$

Similarly, the bivariate empirical repartition H_e of a couple of random variables (X, Y) is equal to

$$H_e(x, y) = \frac{\#(x_i, y_i) \mid x_i \leq x \ and \ y_i \leq y}{N}.$$

We now assume that the random variable X has a distribution function F and that the random variable Y has a distribution function G. The bidimensional copula C of (X, Y) is the cumulative distribution function of the marginals F and G, thus we can say that from the sample $(x_i, y_i)_{i=1,\cdots,N}$, the empirical copula C_e of (X, Y) is equal to nal copula C of (X, Y) is the cumulative distribution function of the marginals F

41

and G, thus we can say that from the sample $(x_i, y_i)_{i=1,\cdots,N}$, the empirical copula C_e of (X, Y) is equal to

$$C_e(u, v) = \frac{\#(x_i, y_i) \mid F(x_i) \leq u \ and \ G(y_j) \leq v}{N}.$$

So that we can give the definition of an empirical copula function:

Definition Let $(x_k, y_k)_{k=1,\cdots,N}$ be a sample of a bivariate random variable. The empirical copula is the function given by

$$C_e\left(\frac{i}{N}, \frac{j}{N}\right) = \frac{\#(x, y) \mid x \leq x_{(i)} \ and \ y \leq y_{(j)}}{N},$$

with $x_{(i)}$ and $y_{(i)}$ the order statistics from the sample[9].

2.3   Correlation measurement

In this section, we will describe what measures of concordance are, or more particularly how we can link them with the copula functions. We will mainly use Roncalli's work [22] to do so. The Kendall's τ will be used in part 4.4 in order to estimate the parameter of a copula function from empirical data.

---

[9]the i-th order statistic is the i-th smallest value of a statistical sample

2.3.1 Concordance

Informally, the concordance of a pair of random variables measures if 'large' values of one are associated with 'large' values of the other, and 'small' values of one with 'small' values of the other.

To be more precise, let (x_i, y_i) and (x_j, y_j) denote two observations from a vector (X, Y) of continuous random variables. We say that (x_i, y_i) and (x_j, y_j) are concordant if $x_i < x_j$ and $y_i < y_j$, or if $x_i > x_j$ and $y_i > y_j$. Similarly, We say that (x_i, y_i) and (x_j, y_j) are discordant if $x_i < x_j$ and $y_i > y_j$, or if $x_i > x_j$ and $y_i < y_j$.

Note the alternate formulation: (x_i, y_i) and (x_j, y_j) are concordant if $(x_i - x_j)(y_i - y_j) > 0$, and discordant if $(x_i - x_j)(y_i - y_j) < 0$.

Using this concept of concordance, we can now define measures of concordance like the Kendall's τ or the Spearman ρ.

2.3.2 Kendall's Tau

The sample version of the quantity known as Kendall's tau is defined in terms of concordance as follows: Let $\{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$ denote a random sample of N observations from a vector (X, Y) of continuous random variables. There are $\binom{N}{2} = \frac{N(N-1)}{2}$ distinct pairs (x_i, y_i) and (x_j, y_j) of observations in the sample, and each pair is either concordant or discordant. Let c denote the number of concordant pairs, and d the number of discordant pairs. Then Kendall's tau for the sample is defined as

$$\tau = \frac{c - d}{c + d} = \frac{2(c - d)}{N(N - 1)}. \tag{19}$$

The population version of Kendall's tau is defined as the probability of concordance minus the probability of discordance

$$\tau = \mathbb{P}\left[(X_1 - X_2)(Y_1 - Y_2) > 0\right] - \mathbb{P}\left[(X_1 - X_2)(Y_1 - Y_2) < 0\right]. \tag{20}$$

For a copula function, Nelsen [21] shows the following equality:

$$\tau = 4 \int \int_{I^2} C(u, v) dC(u, v) - 1. \tag{21}$$

As a concordance measure, the Kendall's τ's range is from -1 to 1. The meaning of a Kendall's τ (between n random variables) which is equal to 1 is that all the data are perfectly concordant, whereas if the Kendall's τ is equal to -1, then all the data are perfectly discordant. A value of the Kendall's τ means that we cannot extract any concordance or discordance from the data.

The dependence measure of Kendall and Spermann can be easily extended to finite families of random vectors whose dimensions are greater than 2. When using those measures, we have two choices. Either we use $2^n - 1$ measures in order to take into account all the random variables, or we use a unique measure. This is the choice which will be made in our study. First, let us generalize the Kendall's τ of a bidimensional copula function to a multidimensional copula functionnal copula function

$$\tau(C_n) = \frac{1}{2^{n-1} - 1}(2^n \int C_n(F_1(x_1), \cdots, F_n(x_n)) dC_n(F_1(x_1), \cdots, F_n(x_n)) - 1)$$
$$= \frac{1}{2^{n-1} - 1}(2^n \int C_n(u_1, \cdots, u_n) dC_n(u_1, \cdots, u_n) - 1).$$

44

As before, we will study the case of 3-dimensional Archimedean copula function.

Let T be the generalized Kendall's τ estimated from the N realizations of 3 random variables U, V and W with the associated copula C_{β_1, β_2}. T is calculated as the average coefficient of (U, V), (V, W) and (U, W):

$$T = \frac{1}{3}(\tau_{emp}(U, V) + \tau_{emp}(V, W) + \tau_{emp}(U, W)).$$

Thus estimating β_1 and β_2 means finding two parameters $\hat{\beta}_1$ and $\hat{\beta}_2$ such as

$$\tau(C_{\hat{\beta}_1, \hat{\beta}_2}) = T.$$

This equation can't be solved directly because both $\hat{\beta}_1$ and $\hat{\beta}_2$ are unknown. However, $\hat{\beta}_1$ can be interpreted as the coefficient of association of U and V with the copula C_{β_1} so that using the section concerning the estimation of the parameter of a 2-dimensional copula function, we can deduce that a semi-parametric estimator $\hat{\beta}_1$ of β_1 is:

$$\hat{\beta}_1 = \tau^{-1}(\tau_{Emp}(U, V)),$$

with $\tau_{Emp}(UV)$ the Kendall coefficient between U and V estimated from the realizations of those two vectors. Thus, the estimator $\hat{\beta}_2$ of β_2 is solution of:

$$\tau(C_{\hat{\beta}_1, \beta_2}) = T.$$

2.3.3 Spearman's Rho

As with Kendall's tau, the population version of the measure of correlation known as Spearman's rho is based on concordance and discordance. To obtain the population version of this measure, we now let (X_1, Y_1), (X_2, Y_2), and (X_3, Y_3) be three independent random vectors. The population version of the Spearman's rho is defined to be proportional to the probability of concordance minus the probability of discordance for the two vectors (X_1, Y_1) and (X_2, Y_3)

$$\rho = 3(\mathbb{P}\left[(X_1 - X_2)(Y_1 - Y_3) > 0\right] - \mathbb{P}\left[(X_1 - X_2)(Y_1 - Y_3) < 0\right]) \tag{22}$$

$$\rho = 12 \int \int_{I^2} [C(u, v) - uv] \, du dv. \tag{23}$$

As a concordance measure, the Spearman's ρ's range is from -1 to 1. The meaning of a Spearman's ρ (between n random variables) which equals 1 is that all the data are perfectly concordant, whereas if the Spearman's ρ equals -1, then all the data are perfectly discordant. A value of 0 of the Spearman's ρ means that we cannot extract any concordance or discordance from the data.

2.3.4 Application

These dependence parameters are of great interest because they enable us to link the correlation coefficients of different types of copula functions. Indeed, the most popular copula functions have a closed formula for the Kendall's tau and the Spearman's rho. Thus, we can compare the dependence measures of different copula functions. For instance, the Student copula and the Gaussian copula have the same Kendall's tau which is equal to $\tau = \frac{2}{\pi} arcsin(\rho)$. The utilization of these dependence measures

can be the measurement of the dependence between two markets, or between two stocks. Besides, it is very interesting to understand that these dependence measures are the translation of the dependence structure implied by the copula function.

3 Archimedean Copula Functions

In this section, we will more particularly focus on the Archimedean copula functions for three reasons:

- Archimedeans copula functions are computationally efficient to implement, generally having closed form solutions,

- they are various different kinds of Archimedeans copula functions,

- and Archimedean copula functions have many interesting properties that we will describe below.

In order to understand more easily the properties of multivariate Archimedean copula functions, we will first introduce the bivariate Archimedean copula functions by showing their properties. Then we will describe algorithms to estimate those copula functions. Finally, we will introduce a method which will be applied in chapter 4.4, which makes it possible to choose a copula function given a data-set.

3.1 2 dimensional (or bivariate) Archimedean copula functions

In this subsection, we will introduce the most important theorems used to describe the Archimedean copula functions. All these theorems have been proven by Nelsen in his book (Nelsen [21]) such as by Genest and MacKay [11], and by Roncalli [22]. Theorems and Definition 3.1.a to 3.1.d will be used to define what Archimedean

copula functions are. Then, Theorems 3.1.e and 3.1.f will show properties of the Archimedean copula functions. Finally, we will show how to estimate a 2-dimensional Archimedean copula function.

Definition 3.1.a Let φ be a continuous and strictly decreasing function from I to $[0, \infty]$ such that $\varphi(1) = 0$. Let φ^{-1}, which will be called the pseudo-inverse of φ be defined on the domain $[0, \infty]$ with $Ran\varphi^{-1} = I$ and

$$\varphi^{-1}(t) = \left[\begin{array}{l} \varphi^{-1}(t), \ \ 0 \leq t \leq \varphi(0) \\ \\ \\ 0, \ \ \varphi(0) \leq t \leq \infty \end{array} \right. .$$

We will see later that Archimedean copula functions are defined by $C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v))$

Theorem 3.1.b Let φ be a continuous and strictly decreasing function from I to $[0, \infty]$, such as $\varphi(1) = 0$. Let φ^{-1} be the pseudo-inverse of φ. Let C be the function defined from I^2 to I by:

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

Then,

$$C(u, 0) = C(0, v) = 0, \qquad \forall u, v \in I^2$$

and

$$C(u, 1) = u \qquad and \qquad C(1, v) = v.$$

Proof

We have:

$$C(u, 0) = \varphi^{-1}(\varphi(u) + \varphi(0)) = 0$$

and

$$C(u, 1) = \varphi^{-1}(\varphi(u) + \varphi(1)) = \varphi^{-1}(\varphi(u)) = u.$$

By symmetry, we have $C(0, v) = 0$ and $C(1, v) = v$.

Theorem 3.1.c Let φ, φ^{-1} and C be such that φ is a continuous and strictly decreasing function from I to $[0, \infty]$, such as $\varphi(1) = 0$, φ^{-1} is the pseudo-inverse of φ and C is defined as in the theorem before. Then C is 2-increasing if and only if, $\forall u_1 \leq u_2$,

$$C(u_2, v) - C(u_1, v) \leq u_2 - u_1 \qquad \forall v \in I. \tag{24}$$

Proof

If C is two-increasing, it is obvious that (24) is verified. On the other hand, if we assume that (24) is verified, let's choose v_1 and v_2 in I such that $v_1 \leq v_2$. Thus, we have from Theorem 3.1.b

$$C(0, v_2) = 0 \leq v_1 \leq v_2 = C(1, v_2).$$

Moreover, as φ and φ^{-1} are continuous, C is also continuous. As a consequence, we can find t in I such that $C(t, v_2) = v_1$. Thus, from Definition 3.1.a,

$$
\begin{aligned}
C(u_2, v_1) - C(u_1, v_1) &= \varphi^{-1}(\varphi(u_2) - \varphi(v_1)) - \varphi^{-1}(\varphi(u_1) - \varphi(v_1)) \\
&= \varphi^{-1}(\varphi(u_2) + \varphi(v_2) + \varphi(t)) - \varphi^{-1}(\varphi(u_1) + \varphi(v_2) + \varphi(t)) \\
&= C(C(u_2, v_2), t) - C(C(u_1, v_2), t) \\
&\leq C(u_2, v_2) - C(u_1, v_2).
\end{aligned}
$$

So that C is 2-increasing.

We can now define copula functions using the functions φ we have defined before.

Theorem 3.1.d Let φ be a continuous, strictly decreasing function from I to $[0, \infty]$, such that $\varphi(1) = 0$ and φ^{-1} is the pseudo-inverse of φ. Then C is a copula function if and only if φ is convex. Such a copula is called an Archimedean copula function, and φ is the generator of this copula function.

Proof

Before going through the demonstration, we should first recall the definition a convex function:

Let f be a continuous function of $[0, +\infty]$ and let s and t be in $[0, +\infty]$, such that $0 \leq s < t$. Then, f is said to be convex if and only if:

$$
f(\frac{s+t}{2}) \leq \frac{f(s) + f(t)}{2}. \tag{25}
$$

We will first assume that C is a copula, and we will demonstrate in Theorem 3.1.c that φ is convex:

As C is a copula, we have demonstrated that

$$C(u_2, v) - C(u_1, v) \leq u_2 - u_1, \qquad \forall v \in I.$$

Thus, if we set $a = \varphi(u_1)$, $b = \varphi(u_1)$, and $c = \varphi(v)$, we obtain

$$\varphi^{-1}(a) + \varphi^{-1}(b + c) \leq \varphi^{-1}(b) + \varphi^{-1}(a + c).$$

Thus, if we now set $a = \frac{s+t}{2}$, $b = s$, $c = \frac{t-s}{2}$ and $\varphi^{-1} = f$, we obtain directly (25).

In the other direction, we will now assume that φ^{-1} is convex. We can now use the same reasoning as before going backwards.

Theorem 3.1.e Let C be an Archimedean copula function with φ its generator.

- C is symmetric, $C(u, v) = C(v, u)$, $\qquad \forall u, v \in I$

- C is associative, $C(C(u, v), w) = C(u, C(v, w)), \forall u, v, w \in I$

- if $c > 0$ is constant, $c\varphi$ is also a generator.

Proof

The proof of this theorem is straightforward, as we only need to write the definition 3.1.a of an Archimedean copula function. For the first point, we have $C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)) = \varphi^{-1}(\varphi(v) + \varphi(u)) = C(v, u)$.

For the second point: $C(C(u,v),w) = \varphi^{-1}(\varphi(\varphi^{-1}(\varphi(u) + \varphi(v))) + \varphi(w)) = \varphi^{-1}(\varphi(u) + \varphi(v) + \varphi(w)) = \varphi^{-1}(\varphi(u) + \varphi(\varphi^{-1}(\varphi(u) + \varphi(w)))) = C(u, C(v,w))$.

And finally for the third point: $C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v)) = 1/c\varphi^{-1}(c\varphi(u) + c\varphi(v)) = \varphi^{-1}(\varphi(u) + \varphi(v))$.

Theorem 3.1.f A copula function C is Archimedean if it has 2 partial derivatives and if there exists an integrable function f from $[0,1]$ to $[0,\infty]$ such as:

$$f(u)\frac{\partial}{\partial v}C(u,v) = f(v)\frac{\partial}{\partial u}C(u,v), \qquad \forall 0 \le u, v \le 1.$$

With $f = \varphi' + c$ with c a constant.

Proof

We have seen earlier that φ is a convex function. As a consequence φ^{-1} exists almost everywhere and thus the partial derivatives $\frac{\partial}{\partial u}C(u,v)$ and $\frac{\partial}{\partial v}C(u,v)$ exist almost everywhere. From

$$\varphi(C(u,v)) = \varphi(u) + \varphi(v)$$

we can deduce that

$$\varphi'(C(u,v))\frac{\partial}{\partial u}C(u,v) = \varphi'(u)$$

and

$$\varphi'(C(u,v))\frac{\partial}{\partial v}C(u,v) = \varphi'(v)$$

53

Moreover, since φ is strictly decreasing, $\varphi'(t) \neq 0$ wherever it exists. Thus we can deduce the result

$$f(u)\frac{\partial}{\partial v}C(u,v) = f(v)\frac{\partial}{\partial u}C(u,v), \qquad \forall 0 \leq u, v \leq 1.$$

Using the definition of a bidimensional copula function, we can now have a method to simulate a random vector (X, Y) of a copula C whose generator is φ.

Let U and S be two independent random variables obtained from an uniform distribution in $[0, 1]$, $X = S$ and Y a random variable. Let $Z = C(X, Y)$. The cumulative distribution function of Z knowing X is:

$$\begin{aligned}
\mathbb{P}(Z \leq z | X = x) &= \mathbb{P}(C(X, Y) \leq z | X = x) \\
&= \mathbb{P}(\varphi^{-1}(\varphi(X) + \varphi(Y)) \leq z | X = x) \\
&= \mathbb{P}(Y \leq \varphi^{-1}(\varphi(z) - \varphi(X)) | X = x) \\
&= \lim_{t \to 0} \mathbb{P}(Y \leq \varphi^{-1}(\varphi(z) - \varphi(X)) | X \in [x - t, x + t])).
\end{aligned}$$

Let y be an outcome of Y. If y is such that $y = \varphi^{-1}(\varphi(z) - \varphi(X))$, we have

$$\begin{aligned}
\mathbb{P}(Z \leq z | X = x) &= \lim_{t \to 0} \frac{\mathbb{P}(Y \leq y, X \leq x + t) - \mathbb{P}(Y \leq y, X \leq x - t)}{\mathbb{P}(X \in [x - t, x + t])} \\
&= \frac{\partial H(x, y)}{\partial x} \\
&= \frac{\varphi'(x)}{\varphi'(z)}.
\end{aligned}$$

Let U be defined by $U(x) = \mathbb{P}(Z \leq x | X = x)$ and $x \in [0, 1]$, U is then a random variable obtained from an uniform distribution in $[0, 1]$. Thus, we have

$\frac{\varphi'(x)}{\varphi'(z)} = U$ which implies that $Z = \varphi'^{-1}(\frac{\varphi'(X)}{U})$ and $Y = \varphi^{-1}(\varphi(Z) - \varphi(X))$, so that $\mathbb{P}(X \leq x, Y \leq y) = C(x, y)$.

We can summarize this demonstration with these schemes in three steps:

- Step 1: Generate two independent random variables S and U from an uniform distribution in $[0, 1]$.

- Step 2: Calculate $Z = \varphi'^{-1}(\varphi'(S)/U)$.

- Step 3: $X = S$ and $Y = \varphi^{-1}(\varphi(Z) - \varphi(X))$.

3.2 Examples of Archimedeans copula functions

In the following, we will present several families of Archimedean copula functions, and their main properties. Those copula functions can be found for example in Bouye's article [3]. The copula functions we will present in this section will be applied in the fourth chapter in order to see their empirical properties and determine which one fits better to an empirical distribution.

3.2.1 Clayton copula functions

$$C_C(u, v, \theta) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}},$$

with $\theta \geq 0$.

This copula has heavy concentration of probabilities near (0,0) so it correlates small losses. The fact that this copula has no tail dependence made it very similar to the Gaussian copula.

The generator of this copula is

$$\varphi(t) = t^{-\theta} - 1$$

and its Kendall tau is equal to

$$\tau = \frac{\theta}{\theta + 2}.$$

We can check that:

$$C_C(u, v, \theta) = (\varphi(u) + \varphi(v) + 1)^{\frac{1}{\theta}}.$$

Moreover,

$$t = (\varphi(t) + 1)^{-\frac{1}{\theta}}.$$

Thus,

$$\varphi^{-1}(x) = (x + 1)^{-\frac{1}{\theta}},$$

and

$$C_C(u, v, \theta) = \varphi^{-1}(\varphi(u) + \varphi(v)).$$

Secondly, we can also check the formula of the Kendall's τ, using Genest and MacKay [11] who have demonstrated that:

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(u)}{\varphi'(u)} du.$$

Thus,

$$\tau = 1 + 4 \int_0^1 \frac{1 - u^{-\theta}}{\theta u^{-\theta-1}} du$$

$$= 1 + \frac{4}{\theta} \int_0^1 (u^{\theta+1} - u) du$$

$$= 1 + \frac{4}{\theta} \left[\frac{u^{\theta+2}}{\theta + 2} - \frac{u^2}{2} \right]_0^1$$

$$= \frac{\theta}{\theta + 2}.$$

3.2.2   Frank copula functions

This copula is characterized by upper and lower tail independence. The Frank copula functions family is defined for $\beta \neq 1$ and strictly greater than 0 by:

$$C_\beta(u, v) = \frac{\ln(1 + \frac{(\beta^u - 1)(\beta^v - 1)}{\beta - 1})}{\ln(\beta)}.$$

And the generator φ_β is:

$$\varphi_\beta(t) = -\ln(\frac{1 - e^{-\beta t}}{1 - e^{-\beta}}).$$

Finally, the Kendall's tau of Frank copula is equal to:

$$\tau = 1 - \frac{4}{\beta} \left(1 - \int_0^\beta \frac{t}{e^t - 1} dt \right).$$

3.2.3 Gumbel copula functions

$$C_G(u, v, \delta) = e^{-\left((-\ln(u))^\delta + (-\ln(v))^\delta\right)^{\frac{1}{\delta}}},$$

with $\delta \geq 1$.

This copula has more probability concentrated in the tails than does Frank's. It is also asymmetric, with more weight in the right tail. Its main properties are that its lower tail dependence is equal to zero whereas its upper tail dependence is equal to $2 - 2^{-\frac{1}{\delta}}$. The Kendall's tau of the Gumbel copula is equal to $\tau = 1 - \frac{1}{\delta}$.

Finally the generator of this copula function is equal to:

$$\varphi(t) = (-\ln(t))^\delta.$$

3.3 Estimation of Archimedeans copula functions

In the case of a two variable Archimedean copula function, the copula function is entirely known as soon as the parameter is known, and the family of the copula is chosen. We will now study several methods to estimate the copula.

3.3.1 Semi-parametric estimation of an Archimedean copula function

As we have seen in the introduction, we will mainly consider in this thesis the problem of the semi-parametric estimation of a copula function. This semi-parametric estimation will be mainly based on the empirical copula function presented in 2.2.4. The main advantage of this estimation is that it is not necessary to estimate the marginal of the copula function. Moreover, the estimation of the empirical copula function is relatively easy. However, the empirical copula has no closed formula. That is why a second step of our study will be to find a parametric Archimedean copula function whose dependence structure will be as close as possible to our empirical copula function. In this section, we will first present the main theorems which will then be used in 3.4 in order to describe the algorithm which will then be applied in 4.4.

This semi-parametric estimation has been proposed by Genest and Rivest [12]. The main idea is that the copula $C_\varphi(x, y) = \varphi^{-1}(\varphi(x) + \varphi(y))$ is uniquely determined by the function $K(v) = v - \frac{\varphi(v)}{\varphi'(v)}$. To prove this idea, we will use the following theorem:

The Proof of the following theorem can be found in Genest and Rivest [12]

Theorem 3.3.1.a *Let X and Y be two uniform random variables, and the copula function $C(x, y)$ defined by $C(x, y) = \varphi^{-1}(\varphi(x) + \varphi(y))$, with φ convex and decreasing on $[0, 1]$ with $\varphi(1) = 0$.*

Let $U = \frac{\varphi(X)}{\varphi(X) + \varphi(Y)}$, $V = C(X, Y)$ and $\lambda(v) = \frac{\varphi(v)}{\varphi'(v)}$ for $0 < v \leq 1$.

Then

1. *U is uniformly distributed on [0,1],*

2. V *is distributed with respect to the law* $K(v) = v - \lambda(v)$ *on* $(0, 1)$ *and*

3. U *and* V *are independent random variables.*

Thus, we can estimate φ by solving the differential equation:

$$\frac{\varphi(v)}{\varphi'(v)} = v - K(v),$$

which gives:

$$\varphi(v) = exp(\int_{v_0}^{v} \frac{1}{\lambda(t)} dt),$$

with $0 < v_0 < 1$ a constant.

We will discuss in more detail in 3.4 how we can use this fundamental theorem to perform the choice of the best copula function given an empirical dependence structure.

Theorem 3.3.1.b *Let* X *and* Y *two uniform random variables with the respective copula function* $C(x, y)$. *For* $0 \leq v \leq 1$, *let* K *be defined by:* $K(v) = \mathbb{P}(C(X, Y) \leq v)$ *and* $K(v^-)$ *be defined by:* $K(v^-) = lim_{t \to v} K(t)$.

Then, the function $\varphi(v)$ *defined by* $\varphi(v) = exp(\int_{v_0}^{v} \frac{1}{\lambda(t)} dt)$ *is convex, decreasing and satisfies* $\varphi(1) = 0$ *if and only if* $K(v^-) > v$ *for all* $0 < v < 1$.

Using the preceding theorem, we can now determine a method to estimate a copula C with a semi-parametric procedure. Let $\{(X_1, Y_1), \ldots, (X_N, Y_N)\}$ be a sample of random variables obtained from a bivariate law $H(x, y)$ with continuous marginals

$F(x)$ and $G(y)$ and a copula function $C(x, y)$ $(C(F(x), G(y)) = H(x, y))$. We assume that we want to estimate the copula C if C is an Archimedean copula function. The method we will use is independent of the marginals so that we can use uniform marginals and then generalize to any kind of marginals. So that H and C can be mixed up as they are equivalent.

We have seen before that Archimedean copula functions are characterized by the behavior of the stochastic random variables $V = H(X, Y)$, so that to estimate the copula, we can estimate the univariate cumulative distribution function $K(v) = \mathbb{P}(H(X, Y) \leq v) = \mathbb{P}(C(F(X), G(Y) \leq v)$ on $(0, 1)$. A two step method can be derived:

- Step 1: Determine the empirical bivariate cumulative distribution function $H_N((x, y))$ associated with H.

- Step 2: Calculate $H_N(X_i, Y_i)$ for $i = 1, \ldots, N$ and use pseudo-observations to build a 1-dimensional empirical cumulative distribution function for K.

We will explain this method with more details in section 3.4.

3.3.2 Using Kendall's τ or Spearmann's ρ to estimate an Archimedean copula function

When the marginal distributions are unknown, we must use the semi-parametric method we have seen before to estimate an Archimedean copula function. After studying the general concept of the semi-parametric method, we will apply it to the case of the Kendall's τ and the Spearmann's ρ. We have studied before the

correlation measurement and we have seen that those two measures are based on the notion of concordance. Remember that for (X, Y) a pair of continuous random variables of copula C, the Kendall's τ is given by:

$$\tau(X, Y) = 4 \int \int_{[0,1]^2} C(u, v) dC(u, v) - 1,$$

which is equivalent to $\tau(X, Y) = \mathbb{E}(C(U, V)) - 1$.

Moreover, if the copula is an Archimedean copula C_β with a parameter β, then the Kendall's τ can be written as $\tau(C_\beta)$.

As a consequence of its definition, the empirical estimator of τ is given by:

$$\tau_{emp} = \frac{2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} X_{ij} Y_{ij}}{N(N+1)},$$

with $X_{ij} = 1$ if $x_i \leq x_j$, $X_{ij} = -1$ if $x_i > x_j$, $Y_{ij} = 1$ if $y_i \leq y_j$ and $Y_{ij} = -1$ if $y_i > y_j$

Thus, we can now define an estimator of the parameter β of the copula function:

$$\hat{\beta} = \tau^{-1}(\tau_{emp}).$$

Using a result demonstrated by Genest and Mackay [11], we finally obtain that:

$$\tau(C_\beta) = 4 \int_0^1 \frac{\varphi_\beta(t)}{\varphi_\beta'(t)} dt + 1.$$

Similarly, we can show the same methodology using the Spearman's ρ which is defined by:

$$\rho(X, Y) = 12 \int \int_{[0,1]^2} uv \, dC(u, v) - 3.$$

The empirical estimator of the Spearman's ρ is given by

$$\rho_{emp} = 1 - 6 \sum_{i=1}^{N} \frac{D_i^2}{N(N^2 - 1)}.$$

where D_i is the rank difference [10] between x_i and y_i. Finally, we can estimate the parameter of the copula function by:

$$\hat{\beta} = \rho^{-1}(\rho_{emp}).$$

3.3.3 The simulation of a 3-dimensional Archimedean copula functions

In this sub-section, before going further into the applications of the theorems and definitions we have studied before, we will just have a quick look at the method we can use to simulate a 3-dimensional copula function.

In the following, we define the product copula by $\Pi(u, v) = uv = exp(-((-\ln(u)) + (-\ln(v))))$. For an n-dimensional product copula function, we have for $u = (u_1, \ldots, u_n)$

$$\Pi^n(u) = u_1 \cdots u_n = exp(-((-\ln(u_1)) + \ldots + (-\ln(u_n)))).$$

---

[10]To calculate the rank difference between between x_i and y_i, you have to calculate the difference between the order statistics of each sample. Moreover, if two x_i or y_i have the same value, the order statistics is the same for each x_i or y_i and equals to the average of the order statistic

This example is a good illustration of the fact that we can generalize the notion of a copula function from the bivariate case:

$$C^n(u) = \varphi^{-1}(\varphi(u_1) + \ldots + \varphi(u_n)).$$

This notation is called the serial iterates of the bidimensional Archimedean copula function generated by φ. We can state that $C^2(u_1, u_2) = C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2))$. Thus for all $n \geq 3$, we have:nal Archimedean copula function generated by φ. We can state that $C^2(u_1, u_2) = C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2))$. Thus for all $n \geq 3$, we have:

$$C^n(u_1, \ldots, u_n) = C(C^{n-1}(u_1, \ldots, u_{n-1}), u_n).$$

However, this method does not provide a n-dimensional copula function for all generators φ, continuous, strictly decreasing and convex. Thus, we have to provide some additional properties to obtain an Archimedean copula function.

Starting from the bivariate case seen earlier, this Archimedean copula function is obtained recursively:

$$C^n(u_1, \ldots, u_n) = \varphi_n^{-1}(\varphi_n(C_{n-1}(u_1, \ldots, u_{n-1})) + \varphi_n(u_n))$$

with $0 \leq u_1, \ldots, u_n \leq 1$ and the generators φ_i strictly decreasing, continuous and convex.

We can apply this formula to the generation of a copula function with 3 dimensions. Thus, we have 2 functions φ_1 and φ_2, which are dependent on 2 parameters

β_1 and β_2 defining the copula function $C_{\beta_1\beta_2}$. We will generate a random vector (X, Y, Z), whose marginals are uniform on $[0, 1]$, with the copula function $C_{\beta_1\beta_2}$ generated by φ_1 and φ_2:

- Step 1: We generate 3 random variables X, U and T from a uniform distribution on $[0, 1]$,

- Step 2: We calculate $W_1 = (\varphi_1^{-1})'(\varphi_1(X)/U)$.

- Step 3: Let $Y = \varphi_1^{-1}(\varphi_1(W_1) - \varphi_1(X))$.

- Step 4: We calculate $W_2 = F^{-1}(T)$, with F the conditional cumulative distribution function of $W_2 = C(X, Y, Z)$ knowing X and Y.

- Step 5: $Z = \varphi_2^{-1}(\varphi_2(W_2) - \varphi_2(\varphi_1^{-1}(\varphi_1(X) + \varphi_1(Y))))$.

Thus we have seen that Archimedean copula functions are defined from generator functions φ which depends on one or more parameters β_i. The proof of this algorithm can be found in Hillali [13]. We will give the main ideas of this demonstration in the following proof:

Proof

Let X and U be two independent random variables drawn from a uniform distribution on $[0, 1]$.

We will now try to determine the random variable Y such that (X, Y) verifies the same distribution as the copula H which is derived from a continuous, convex and strictly decreasing function φ_1.

Let $W_1 = H(X, Y)$. Then the cumulative distribution function of W_1 given X is:

$$\mathbb{P}(W_1 \leq w | X = x) = \mathbb{P}(H(X, Y) \leq z | X = x)$$
$$= \mathbb{P}(\varphi_1^{-1}(\varphi_1(X) + \varphi_1(Y)) \leq z | X = x)$$
$$= \mathbb{P}(Y \leq \varphi_1^{-1}(\varphi_1(z) - \varphi_1(X)) | X = x)$$
$$= \lim_{t \to 0} \mathbb{P}(Y \leq \varphi_1^{-1}(\varphi_1(z) - \varphi_1(X)) | x - t \leq X \leq x + t).$$

If we define y by $y = \varphi_1^{-1}(\varphi_1(z) - \varphi_1(X))$ a value of Y. Then:

$$\mathbb{P}(W_1 \leq w | X = x) = \lim_{t \to 0} \frac{\mathbb{P}(Y \leq y, X \leq x + t) - \mathbb{P}(Y \leq y, X \leq x - t)}{\mathbb{P}(x - t \leq X \leq x + t)}$$
$$= \frac{\partial H(x, y)}{\partial x}$$
$$= \frac{\varphi_1'(x)}{\varphi_1'(w)}.$$

If we define U by $U(x) = \mathbb{P}(W_1 \leq w | X = x)$ with $x \in [0, 1]$, the U is random variable uniformly distributed on $[0, 1]$. As a consequence:

$$U = \frac{\varphi_1'(X)}{\varphi_1'(W_1)},$$

so that

$$Z = \varphi_1'^{-1}\left(\frac{\varphi_1(X)}{U}\right).$$

Finally, Z has been constructed so that $Y = \varphi_1^{-1}(\varphi_1(Z) - \varphi_1(X))$ and $\mathbb{P}(X \leq x, Y \leq y) = H(x, y)$.

66

We will now demonstrate Steps 4 and 5:

We will use F the cumulative distribution function of $W_2 = C(X, Y, Z)$ given X and Y:

$$F(w_2) = \mathbb{P}(W_2 \leq w_2 | X = x, Y = y)$$
$$= \mathbb{P}(C(X, Y, Z) \leq w_2 | X = x, Y = y)$$
$$= T$$

with $w_2 \in [0, 1]$ and T a uniform random variable in $[0, 1]$.

Thus we have

$$T = \mathbb{P}(\varphi_2^{-1}(\varphi_2(Z) + \varphi_2(\varphi_1^{-1}(\varphi_1(X) + \varphi_1(Y)))) \leq w_2 | X = x, Y = y)$$
$$= \mathbb{P}(Z \leq \varphi_2^{-1}(\varphi_2(w_2) - \varphi_2(\varphi_1^{-1}(\varphi_1(X) + \varphi_1(Y)))) \leq w_2 | X = x, Y = y).$$

Moreover, as X, Y and Z are three continuous $\mathbb{R}$-random variables, we can describe the law of Z as a limit with $t \to 0$ of:

$$T = \frac{\mathbb{P}(Z \leq \varphi_2^{-1}(\varphi_2(w_2) - \varphi_2(\varphi_1^{-1}(\varphi_1(X) + \varphi_1(Y)))) \leq w_2 | x \leq X \leq x + t, y \leq Y \leq y + t)}{\mathbb{P}(x \leq X \leq x + t, y \leq Y \leq y + t)}.$$

Let H_1 be the 2-dimensional cumulative distribution function of (X, Y). We denote respectively by $f(x), g(y)$ and $h_1(x, y)$ the density functions of X, Y and H_1. Then:

$$T = \frac{\frac{\partial^2 C}{\partial x \partial y} - \frac{\partial C}{\partial x} - \frac{\partial C}{\partial y}}{h_1(x, y) - f(x) - g(y)}.$$

67

In order to simplify the calculus, we will simulate 3 random variables X, Y and Z from a 3 dimensional Frank copula whose parameters are γ_1 and γ_2 associated respectively to the generators φ_1 and φ_2. We will use for the calculations: $u = \varphi_1^{-1}(\varphi_1(x) + \varphi_1(y))$, $A_1 = \varphi_1'(x) + \varphi_1'(y)$, $A_2 = \varphi_1'(x)\varphi_1'(y)$, $B = \varphi_1'(u)$, $C = \varphi_2'(u)$, $D = \varphi_2''(u)$, $E = \varphi_1''(u)$

$$F = \frac{A_1 \gamma_1^u + A_2}{B \gamma_1^u};$$

$G = CA_1 B^2$, $M = DBA_2$, $I = A_2 CE$, $J = A_2 C^2 B$, $K = B^3 F$, $L = K \log(\gamma_2)$.

Then, we have:

$$F(w_2) = T = \frac{(\gamma_2^{w_2}(G - M + I) + J)(1 - \gamma_2)}{\gamma_2^{2w_2} L}.$$

Using the inverse cumulative distribution function, we get W_2, and by construction of H, we find Z.

3.4 Application to the choice of an Archimedean copula function [4]

In the previous sections, after having made a general presentation of copula functions, we have seen methods to estimate the parameters of those copula functions. Besides, we have seen that the family of copula functions is a very large family. We have focused our attention on the family of Archimedean copula function, and we have seen that each copula function has its own properties and fits a different dependence structure. As a consequence, we now have to focus on the method to try to choose

the best copula by comparing the dependence structure of each copula function to our data-set.

As we have seen before, we can link the parameter of a copula function with the concordance measure

$$\tau = f(\theta).$$

In our approach, τ is observed and f is a function dependent on the choice of the copula. For example, for a Gumbel copula, $\tau = \frac{\theta-1}{\theta}$. As a consequence, it is possible to estimate θ as

$$\theta = f^{-1}(\tau).$$

The first step of our process is thus to estimate the parameter θ based on the observation of τ. This estimation will be made for each kind of copula.

Then, we will have to make a choice between all those copula to find the copula which best fits our distribution, and describe accurately the dependence structure between the data. Intuitively, we will choose the copula type which lies the closest to the empirical dependence structure. Namely, the optimal copula is the function which minimizes the observed errors relatively to the empirical copula function.

The principle for selecting the optimal copula is simple. For this purpose, we will introduce a discrete L^2 norm which will measure the distance between a theoretical copula C which will belong to our copula set $\bar{C}$ and the empirical copula $\hat{C}$ that will be estimated on the observed data. Thus, we will obtain the optimal copula C^* that

is to say the copula which gives the best description of the dependence structure of our sample of data.

To do so, we will first introduce the distance $\hat{d}_2$ of the discrete L^2 norm:

$$\hat{d}_2(C,\hat{C}) = \left(\sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \left[C(\frac{t_1}{T}, \frac{t_2}{T}) \hat{C}(\frac{t_1}{T}, \frac{t_2}{T}) \right]^2 \right)^{1/2} ,$$

where C belongs to $\bar{C}$, and T corresponds to the number of observations. Therefore, the optimal copula function C^* describing the dependence structures we study, given our copula set $\bar{C}$, has to satisfy:

$$C^* = argmin_{C \in \bar{C}} \; \hat{d}_2(C,\hat{C}) \; .$$

Another method can be used to select the optimal copula function among a set of copula functions. This method has been first described by Genest and Rivest [11] and it is based on the observation of an unobserved random variable $Z_i = F(X_{1i}, X_{2i})$ that has a distribution function $K(z) = \mathbb{P}(Z_i \leq z)$. Genest and Rivest showed that this distribution function is related to the generator of an Archimedean copula function through the expression:

$$K(z) = z - \frac{\varphi(z)}{\varphi'(z)}.$$

The identification of φ is thus made in three steps:

1. Estimate the Kendall's correlation coefficient.

2. Construct a semi-parametric estimate of K by first determining the pseudo-observation $Z_i = $ number of X_{1j}, X_{2j} such that $X_{1j} < X_{1i}$ and $X_{2j} < X_{2i}$ for $i = 1, \cdots, n$. And then construct the estimate of K as $K_n(z) =$ proportion of $Z_i \leq z$.

3. Construct a parametric estimate of K using the relationship $K_\varphi(z) = z - \frac{\varphi(z)}{\varphi\prime(z)}$.

For example, we can test different types of copula by first estimating $Kendall's \tau$ to calculate an estimate of the parameter of the copula function. We then use this estimate of the copula to further estimate the generator of the copula. Finally, we use this estimate of the generator to estimate $K_{\varphi n}$.

We then repeat Step 3 for different choices of φ. We will finally compare these results with the semi-parametric estimate constructed in step 2, and select the choice of φ so that the parametric estimate $K_{\varphi n}$ most closely resembles (in terms of L^2 norm) the semi-parametric estimates.

This algorithm will be used in part 4.4 in order to model the joint distribution of equity returns by choosing the most appropriate copula function.

4 Application to 1st-to-default Basket CDS Pricing

Before going through the pricing process, let's have a look at a very simple example which will remind us of the reason why studying correlation is a paramount problem in pricing Basket CDS. Let's take the example of a simple Basket CDS, which reference to two bonds. Assume that the default correlation between the two assets is equal to zero, which means that if a company defaults, we cannot make any assumption about the likeliness of default of the other company. Then it is intuitively clear that the probability of one default in the Basket CDS is strictly greater than the probability of two defaults. Thus, the value of the 1st-to-default Basket CDS is greater than the value of 2nd-to-default Basket CDS. Now let's assume that the default correlation between the two bonds in the Basket CDS is equal to one. Then, as soon as a bond defaults, the other one defaults too, and the value of the 1st-to-default Basket CDS and the 2nd-to-default Basket CDS are equal. This very simple example shows us that the study of default correlation is a key point in the valuation of such credit derivatives.

Default correlation is also a time dependent problem. To illustrate this, we can consider the following very intuitive example. Take two companies whose default correlation is not equal to zero. Then we can assert that the probability that both companies default within two years is greater than the probability that both companies default within one year.

In this chapter, our aim is to describe how we can perform the pricing of a very simple Basket CDS. The first subsection gives the process used to price a 1st-to-default Basket CDS. Then, we present the result of this pricing and compare the

price obtained by different dependence structures which are modeled by different copula functions. Finally, we apply another very interesting algorithm described in chapter 3.4 which enables us to choose the best copula function to fit a given data-set.

4.1   The Pricing Process

The pricing process using copula functions is much more simple than the direct use of joint distributions, because it lets us separate the study of the marginal functions (the credit curves), and the dependence between those marginal functions. This way, we can use different copula functions to model different kinds of dependence between the marginal functions. This pricing process has been extensively described by Li [18].

As a copula function based model can be very complicated to fit to the market data, we will not study in this thesis a market based model. Indeed, our portfolio will be made with simplified bonds. The default characteristics of those simplified bonds will not be extracted from market data, but from Moody's table of default probabilities. Thus, this very simple model aims to understand the basic mechanisms implied by the utilization of copula functions.

In this section, we will use a portfolio of 6 credits with a recovery rate equal to zero (i.e. in case of a default, the credit is worth zero). The correlation between each credit is supposed to be equal. However, this assumption is very easy to relax, but using different correlation coefficients will hide the effect of dependence on the price

of the nth-to-default Basket CDS. Finally, the product priced is a contract which pays 1\$in case of the default of one of the credit of the portfolio [11].

We now describe the process to price a 1st-to-default basket CDS. This process uses the Monte-Carlo simulation to simulate a random sample. It means that we will not derive the price of the basket CDS from a closed formula but we will choose random variables which will be used to model the default time. Then, we will use the dependence structure given by the copula function to calculate the price of the basket CDS. We will then do this random choice again (typically several thousand times) and each time calculate a price for the basket CDS. Finally, the average of the prices will converge to the price of the basket CDS.

As a consequence each Monte-Carlo simulation will be split into three main steps:

Model the joint distribution with the copula: In our study, we will use the multivariate normal copula function (cf. 2.2.1). The first step is to simulate $Y_1, Y_2, \cdots, Y_n$ from an n-dimensional normal distribution with correlation coefficient matrix Σ.

Obtain the corresponding marginal distributions: After obtaining the sample $Y_1, Y_2, \cdots, Y_n$, we will use a percentile-to-percentile mapping to obtain the default times $T_1, T_2, \cdots, T_n$ using $T_i = F_i^{-1}(N(Y_i))$ [12].

Calculate the price of the 1st-to-default basket CDS: Knowing the first default time in the portfolio, we can now calculate the price of our derivative.

---

[11]This pricing also uses risk-free interest rate. This interest rate will always be equal to 5% in our applications

[12]F_i will be derived from Moody's historical default times

To conclude, remember that the utilization of copula functions in our study will let us study, in part 4.3, portfolios which have different dependence structures and see that the prices of such basket CDS is a function of the dependence structure.

4.1.1 Model the joint distribution with the copula

A widely used method for drawing a random vector Y from the n-dimensional multivariate normal distribution with mean vector μ (in our study, this vector is equal to zero) and correlation matrix Σ (required to be symmetric and positive definite) works as follows:

1. Compute the Cholesky decomposition (matrix square root) of Σ, that is, find the unique lower triangular matrix A such that $LL^T = \Sigma$.

2. Let $Z = (z_1, \ldots, z_n)$ be a vector whose components are n independent standard normal variates.

3. Let Y be LZ.

4.1.2 Obtain the corresponding marginal distributions

The simulation of the time to default is obtained from the cumulative default probabilities given by Moody's. To obtain T_i, we compare $N(Y_i)$ with Moody's data. However, as the cumulative default probability is not a continuous probability (ie it is only given for discrete time: each time a year), we have to make a choice to calculate the time to default: [13]

---

[13]In this study, we will use the first possibility, and keep it to be consistent

1. We can suppose that if $N(Y_i)$ is less than or equal to the cumulative default probability at time T_i, then we consider that the default occurs at time T_i.

2. We can suppose that if $N(Y_i)$ is less than or equal to the cumulative default probability at time T_i, then we consider that the default occurs at time T_{i-1}.

3. We can suppose that if $N(Y_i)$ is less than or equal to the cumulative default probability at time T_i, we will perform a linear regression of the cumulative default probability between T_i and T_{i-1}. Thus, the default time will be equal to

$$T_i \frac{C_i - Ci - 1}{C_i - N(Y_i)}$$

with C_i, the cumulative default probability for the year i.

4.1.3 Calculate the price of the 1st-to-default basket CDS

To perform the calculation of the price of the Basket CDS, we simply compute the present value of the 1$ payoff and finally make the average over all the Monte-Carlo simulation. In order to compute the present value, we use the result we proved in 1.1.

4.2 Results

Thanks to a VBA program based on the algorithm described, we will be able to get some insight about basket CDS and their parameters. To perform the Monte-Carlo algorithm study, we first study the convergence of this algorithm, for several numbers of simulations, and the precision of those simulations. Then, we study the

dependence of the portfolio to different parameters. The first which will be studied is the correlation coefficient between the credit, which is assumed to be constant and equal between all the credits. The second parameter is the influence of the lifetime of the portfolio. Finally, we also study the influence on the price of the portfolio of an increase of n, the number of defaults before the 1\$ payment.

The portfolio that we study is made of 6 credits: 2 are rated Aaa, 2 are rated Baa1 and the two remaining are rated Caa-C. If no other indication is given, the lifetime of the portfolio is equal to 5 years, and the correlation coefficient is equal to 0.3. This portfolio will be referred in the following as the standard portfolio.
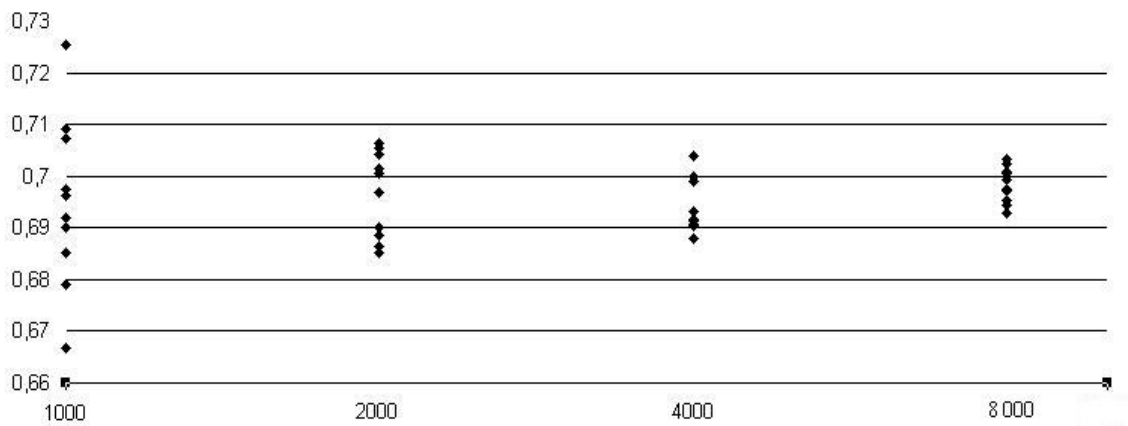

Figure 2: Representation of the price of the 1st-to-default standard Basket CDS as a function of the number of simulations.

Figure 2 shows the convergence of the Monte-Carlo algorithm. As we can see the precision increases with the number of simulations (y-axis), as the incertitude on the 1st-to-default Basket CDS price (x-axis) decreases. So that the average price of $0,6969\$$ (for 10 samples) is obtained, with a standard deviation of $3.4*10^{-3}$.

Another calculation with 100 000 simulations was performed. The result is a mean price of 0.6975\$, with a standard deviation of $1.3 * 10^{-3}$.
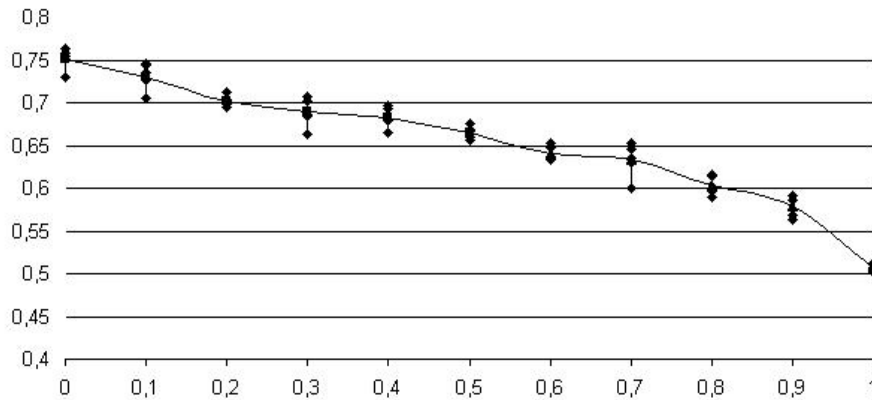

Figure 3: Evolution of the price of the 1st-to-default standard Basket CDS as a function of the correlation coefficient

Figure 3 represents the evolution of the price of the 1st-to-default standard Basket CDS when the correlation coefficient changes. We can see that when the correlation coefficient increases, the price decreases. The reason is that when the correlation coefficient increases, all the entities default times are getting closer. We can consider the limit case of a correlation coefficient equal to 1. Then, it is obvious that if all the bonds have the same rating, they will default at the same time. Thus, when the correlation coefficient increases, the prices of all nth-to-default portfolios tend to become equal. And, as we can see in Figure 4, the price of a nth-to-default Basket CDS tends to decrease as n increases.

Figure 5 represents the evolution of the price of the 1st-to-default Basket CDS when the lifetime of the portfolio increases. We can see that when the lifetime of
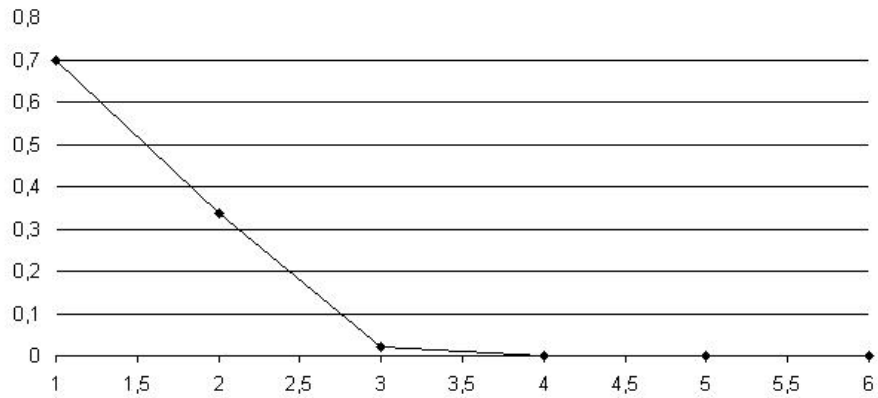
Figure 4: Evolution of the price of the nth-to-default standard Basket CDS as a function of n, the number of defaults before the payment is made
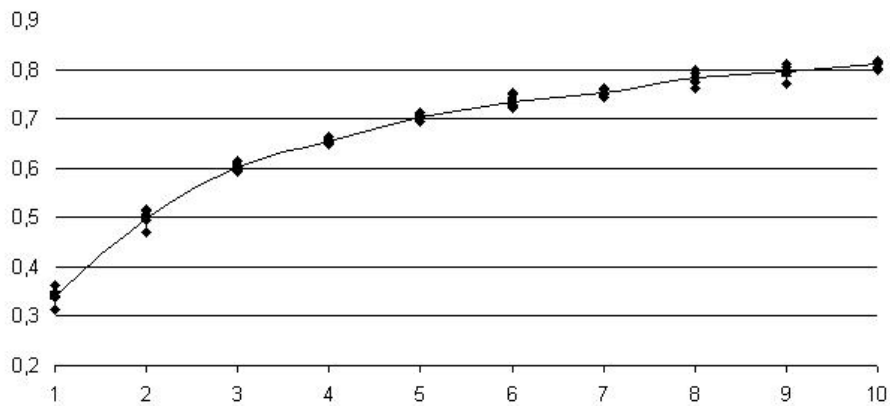

Figure 5: Evolution of the price of the 1st-to-default standard Basket CDS as a function of the lifetime of the portfolio

the portfolio increases, the price also increases, which is consistent with the fact that if the portfolio's lifetime is greater, the probability of a default in the portfolio

intuitively increases.

4.3   Comparison of the different dependence structures

In the previous sections, we have seen how to price a 1st-to-default basket CDS with the Gaussian copula function. We have been able to see that the price of this basket CDS changes with respect to the time until maturity of the portfolio, the rating of the credits included, or the correlation coefficient between those different credits. Another very important parameter which has to be studied when pricing a basket CDS is the dependence structure of the default correlation between the different names of the portfolio. As we have seen before, this dependence structure can be model by different copula functions translating for example the fact that the correlation between different names of a portfolio may increase if the credit spreads increase sharply. As a consequence, we will model a very simple basket CDS in order to show that its price varies when the copula functions are different. In our study, we will model this basket CDS with 6 different copula functions:

- The independent Copula function;

- The perfectly correlated Copula;

- The Gaussian Copula;

- The Gumbel Copula;

- The Frank Copula;

- The Clayton Copula.

Our numerical example will be based on a portfolio of two credits. The payoff will be one dollar if any of the credits defaults. We can assume that the defaults occur for individual assets according to a Poisson process with a deterministic intensity called hazard rate h. As a consequence, the default times T are exponentially distributed with a mean equal to $\frac{1}{h}$. In our example, we will use a Monte-Carlo simulation with 30 000 trials. For each trial, we will draw uniform bivariates from the chosen copula, and then derive the default times from the inverse cumulative exponential distribution, and finally derive the payoff.

Finally, in order to check the results given by our Monte-Carlo simulation, we can derive the analytical solution for the independent case, given the notation explained in section 1.1 (7):

$$V = \frac{h'}{r + h'} * (1 - e^{-t(r+h')}).$$

With $h' = h * n$ in the case of the independent copula function.

In order to be able to compare the different dependence structures, we will use the same Kendall's τ for all the copula functions.

Concerning the Gaussian copula function, we will use the standard bivariate normal distribution with correlation coefficient ρ:

$$C(u, v, \rho) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\Pi\sqrt{1 - \rho^2}} e^{-\frac{-(s^2 - 2\rho*s*t + t^2)}{2(1-\rho^2)}} \, ds \, dt,$$

81

Duration	Perf. Corr	Independent	Gaussian	Clayton	Gumbel	Frank
2 Years	0,178	0,310	0,239	0,211	0,244	0,235
4 Years	0,301	0,455	0,356	0,362	0,383	0,373
6 Years	0,401	0,550	0,438	0,472	0,482	0,474

Table 1: Price of the Basket CDS with h=0,1 and r=0,1

Duration	Perf. Corr	Independent	Gaussian	Clayton	Gumbel	Frank
4 Years	0,0024	0,0021	0,0027	0,0028	0,0030	0,0029

Table 2: Standard Deviation of the price of the Basket CDS with h=0,1 and r=0,1, calculated over 10 times 30 000 simulations

where ϕ and Φ are the univariate standard normal density and cumulative distribution functions, respectively. In our computation, we will use a Taylor expansion for simplicity reasons:

$$C(u, v, \rho) = uv + \rho\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v)).$$

Duration	Independent
2 Years	0,301
4 Years	0,466
6 Years	0,556

Table 3: Analytical Price of the Basket CDS with h=0,1 and r=0,1

Duration	Perf. Corr	Independent	Gaussian	Clayton	Gumbel	Frank
2 Years	0,092	0,192	0,145	0,107	0,138	0,134
4 Years	0,169	0,298	0,229	0,201	0,233	0,225
6 Years	0,232	0,373	0,288	0,276	0,302	0,294

Table 4: Price of the Basket CDS with h=0,05 and r=0,1

In the previous tables, which show the price of our simple basket CDS, we can notice that the price inferred by a different dependence structure can differ greatly from the Gaussian copula. For short maturities, the price of the CDS can double if we consider a perfect correlation or on the contrary no correlation at all. The difference of price between the other copula functions can be as great as 15%. Thus the dependence structure embedded in the portfolio is of great importance for a correct pricing of that kind of product. Moreover, the standard deviation obtained from our simulations shows that the result of simulation has an interesting accuracy.

We can also notice that the price of the basket CDS depending on the copula function chosen always verifies: $P_{Clayton} < P_{Frank} < P_{Gumbel}$. Indeed, the Clayton copula function shows a heavy tail near 0 whereas the Gumbel copula has a heavy right tail. As a greater correlation implies a lower price, the Clayton copula will put more dependence on the default occurring in a near future whereas Gumbel copula will translate the contrary.

Concerning the Frank copula, its structure should be closer to the Gaussian copula. However the Taylor expansion used for simplicity reasons tends to hide this phenomenon. However, for short time until maturity where this expansion is closer

to the real form, we can see that the behavior of the Frank copula and the Gaussian copula are indeed similar.

4.4   How to choose between different dependence structures?

In the preceding sub-section, we compared different dependence structures, and were able to conclude that depending on the dependence structure we choose, the price of the basket CDS will be different. As a consequence, one of the missions of a basket CDS trader will be to determinate which copula function describes the most accurately the portfolio he wants to model. To do so, we will apply the algorithm of a choice of copula to pairs of UK stocks. Indeed, we have seen in 1.4.2 that default correlation between two names of a basket credit derivative can be estimated from equity returns thanks to the model described by Merton [20]. We will thus use the algorithm described in 3.4 to choose the best copula which will describe the dependence structure between two stocks, which is a proxy for the default correlation of those stocks.

We will now apply the algorithm seen in 3.4. The main idea of this algorithm is based on the measurement of the distance between the empirical copula, described by the data, and a copula function (like Frank's copula for instance). The objective is to choose the copula which is closest to the empirical copula, which means that the copula function will describe most accurately the dependence structure between the two time-series we are studying. The goal is then to use this copula function in the pricing of the basket CDS in order to obtain the price which will best fit the
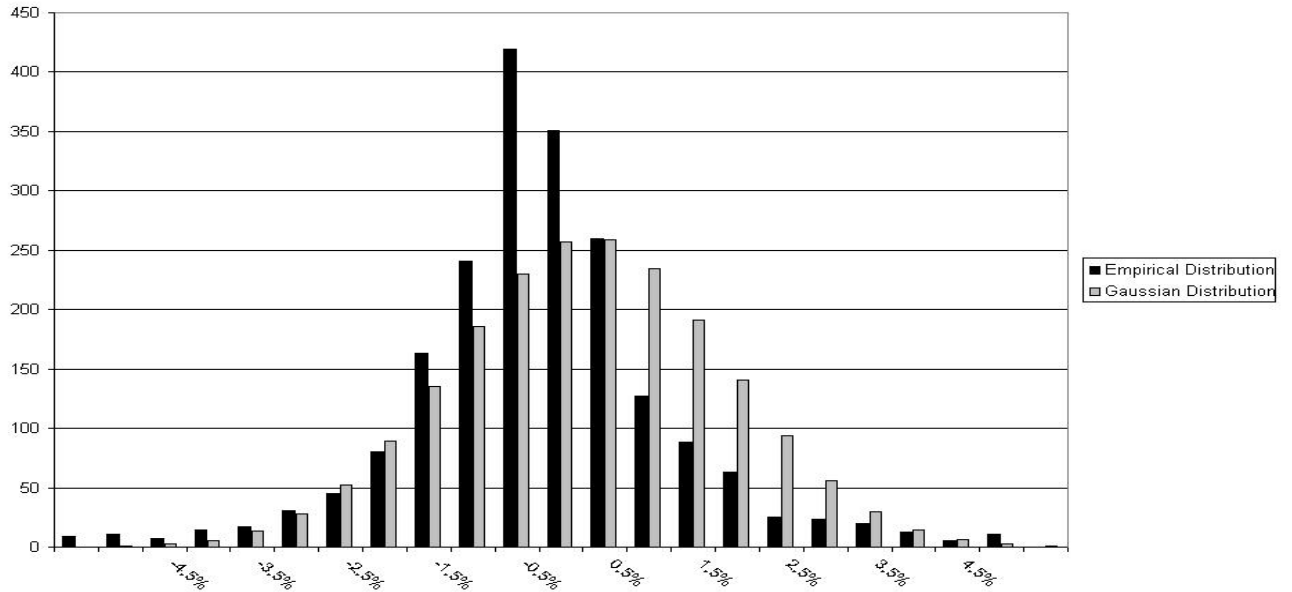
Figure 6: Marginal distribution of HSBC daily returns

dependence structure of the portfolio, and thus the price which will be the closest to the market price.

As our goal is not to draw conclusions on the dependence structure of the financial markets, but just to present a very powerful algorithm to make a choice of copula, we will only focus our study on 2 datasets which will be the daily returns of 3 stocks: HSBC, Royal Bank of Scotland and BP. The daily returns will be taken from May 25th 1999 to May 25th 2007. This dataset represents 1981 daily returns.
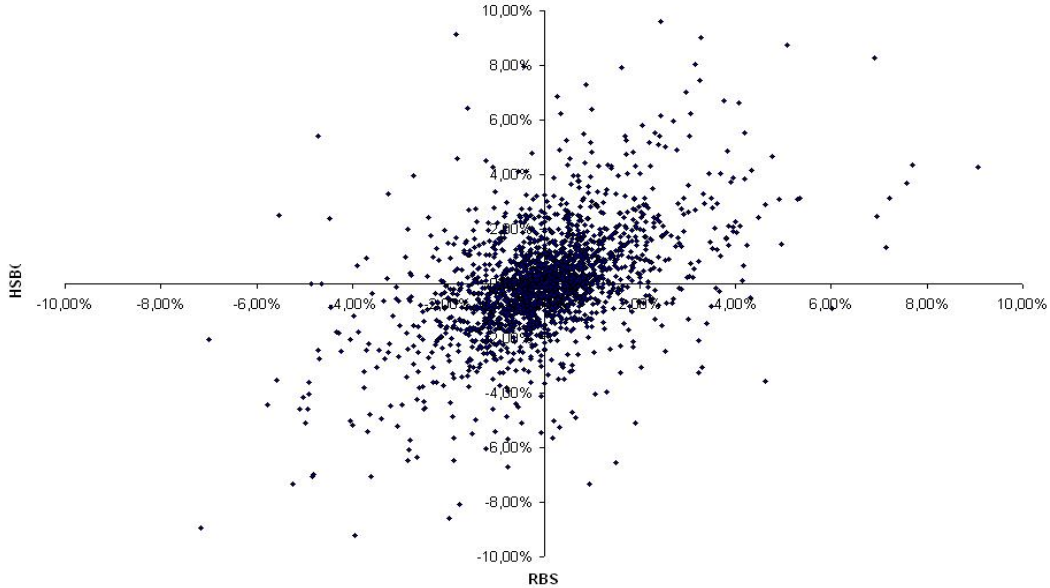
Figure 7: Daily returns of HSBC (x-axis) against RBS (y-axis)

Before focusing on the results of our algorithm, we should first have a very quick look at the structure of our marginal distribution. For example, let's concentrate on HSBC. The daily returns over 8 years have a standard deviation of 1.56%. Moreover, we have represented in figure 6, the distribution of the daily returns of HSBC, compared to the Gaussian distribution. It is very clear looking at that distribution, that the marginal distribution cannot be considered as being Gaussian. Moreover, we have

	Frank Copula	Clayton Copula	Gumbel Copula
Parameter of the copula	1,04	0,63	1,32
Distance	$2,8 \times 10^{-2}$	$9,5 \times 10^{-2}$	$6,2 \times 10^{-2}$

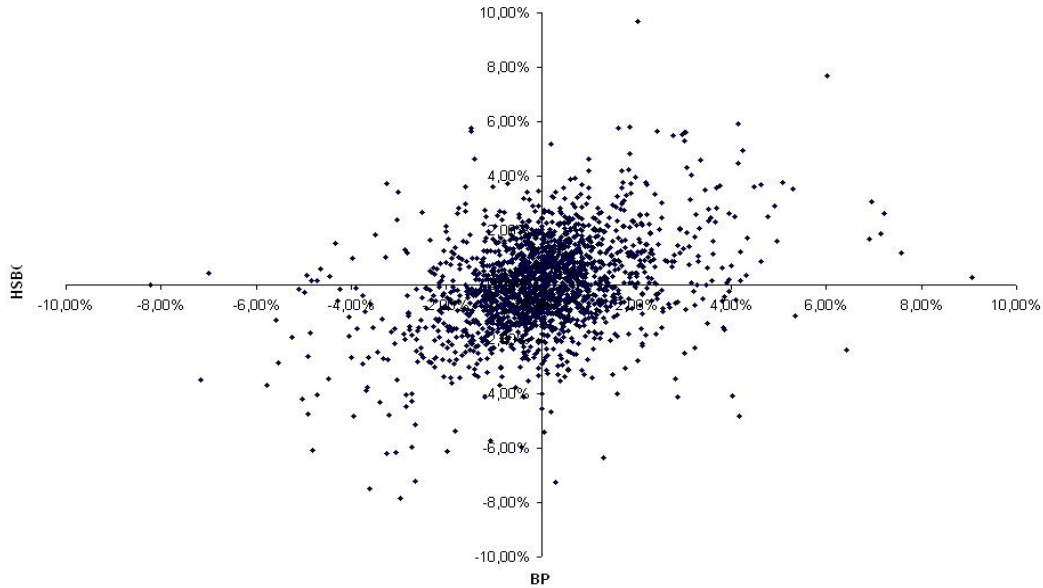Table 5: Distance to the empirical copula for HSBC-BP, Kendall's tau = 0,24

Figure 8: Daily returns of HSBC (x-axis) against BP (y-axis)

calculated the skewness and kurtosis of this distribution which are respectively equal to $-0,06$ and $5,98$, compared to the Gaussian distribution whose skewness equals 0 and kurtosis equals 3. As a consequence, the utilization of the semi-parametric method of estimation for a copula is very accurate, because we don't have to make any hypothesis on the marginal distribution of our dataset. This is one of the most important properties of semi-parametric estimation.

	Frank Copula	Clayton Copula	Gumbel Copula
Parameter of the copula	2,53	1,05	1,52
Distance	$3,7 \times 10^{-2}$	$10,4 \times 10^{-2}$	$6,0 \times 10^{-2}$

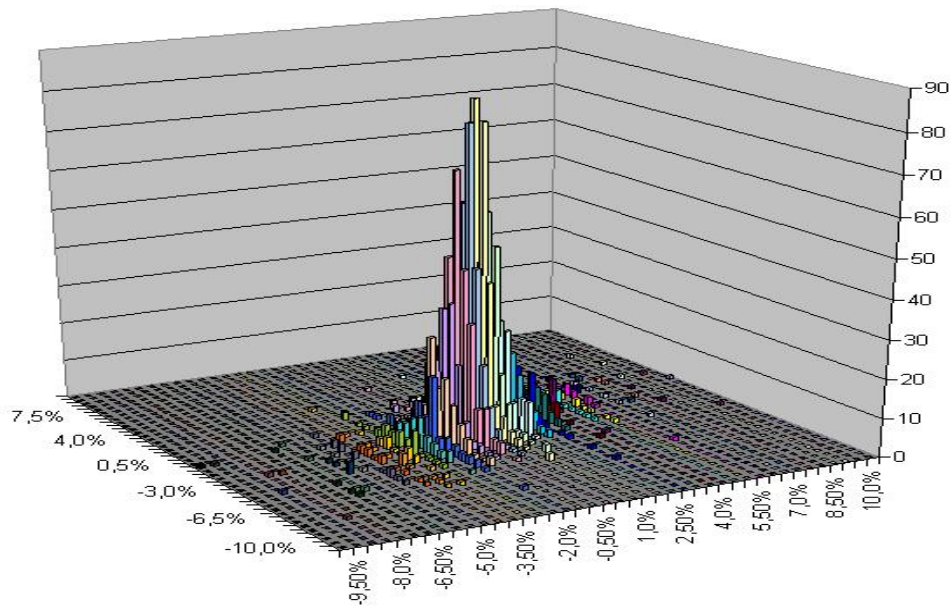Table 6: Distance to the empirical copula for HSBC-RBS, Kendall's tau = 0,34

87

Figure 9: Density of the daily returns(z-axis) of HSBC (x-axis) against RBS (y-axis)

Figure 10: 3-d representation of the empirical copula function for the HSBC-RBS couple

Figure 11: Level curves obtained for theHSBC-RBS couple from different copula function with the same Kendall's tau: from top right to bottom left, the empirical copula, the Gumbel copula, the Clayton copula and the Frank copula

To continue our study, we have drawn 2 graphs: figures 7 and 8 which show the correlation of the daily returns of our two couples of stocks: HSBC and RBS and HSBC and BP. Moreover, figure 9 shows the 3-dimensional density of the daily

returns of HSBC and RBS.

The aim of our study is now to determinate which copula function among a given set of copula function is the best one to fit the empirical market data. For our study, the set of copula functions will be the Frank copula, the Clayton copula and the Gumbel copula. Thus, as we have described in 3.4, we will draw from the market data the empirical copula. Then, we will calculate the Kendall's τ of our dataset which will enable us to determine the parameter of each of our copula functions. Finally, we will calculate $\hat{d}_2$, the distance between our empirical copula and the studied copula. The copula which will be the closest in terms of distance to our empirical copula will be the one describes most accurately the dependence structure between our two stocks.

Even if the goal of this section is to present the result of the algorithm we presented in part 3.4, we will quickly recall the way to derive those results. The input of our program is a dataset given by Reuters which is the closing price of two stocks. From those closing prices, we derive a daily return from which we will calculate the Kendall's τ (see part 2.3.2 for the formula). From this dataset, we will also derive its empirical copula function, using the method described in section 2.2.4. As we described in section 3.4, we will then use Genest and Rivest [12] results exposed in section 3.3.1 in order to calculate the distribution function of the Archimedean copula function. This distribution function (ie the function K) will be represented in figures 12 and 13 for the two examples we study. Finally, we will measure the distance between the distribution function of the empirical copula function and the distribution function of the Archimedean copula function. The copula function which is closest

to the empirical copula function will be the copula function which will describe the dependence structure the most accurately.

As a reminder of the results and details of the algorithm, we will present again its different steps:

1. Estimate the Kendall's correlation coefficient of our dataset.

2. Construct the empirical copula by first determining the pseudo-observation $Z_i = $ number of X_{1j}, X_{2j} such that $X_{1j} < X_{1i}$ and $X_{2j} < X_{2i}$ for $i = 1, \cdots, n$. And then construct the estimate of K as $K_n(z) = $ proportion of $Z_i \leq z$.

3. Construct a parametric estimate of K using the relationship $K_\varphi(z) = z - \frac{\varphi(z)}{\varphi'(z)}$ which will use the $Kendall's \tau$ calculated in the first step. It will be the estimate of the copula we want to model (ie in $\bar{C}$).

4. Finally calculate the distance between the empirical copula and the copula we want to model:

$$\hat{d}_2(C, \hat{C}) = \left(\sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \left[C(\frac{t_1}{T}, \frac{t_2}{T}) \hat{C}(\frac{t_1}{T}, \frac{t_2}{T}) \right]^2 \right)^{1/2}.$$

In tables 5 and 6 we have compiled the results of our study. The parameter of the copula function has been calculated using the empirical value of the $Kendall's \tau$ we have calculated from our data-set, and the formula from chapter 3.2. On the second line, we have presented the distance of the copula function to the empirical copula. In both cases, we can see that the closest copula function to the empirical copula function is Frank copula. This result could be explained by the fact that we have a good correlation for small returns, however, for larger returns, we have a wider
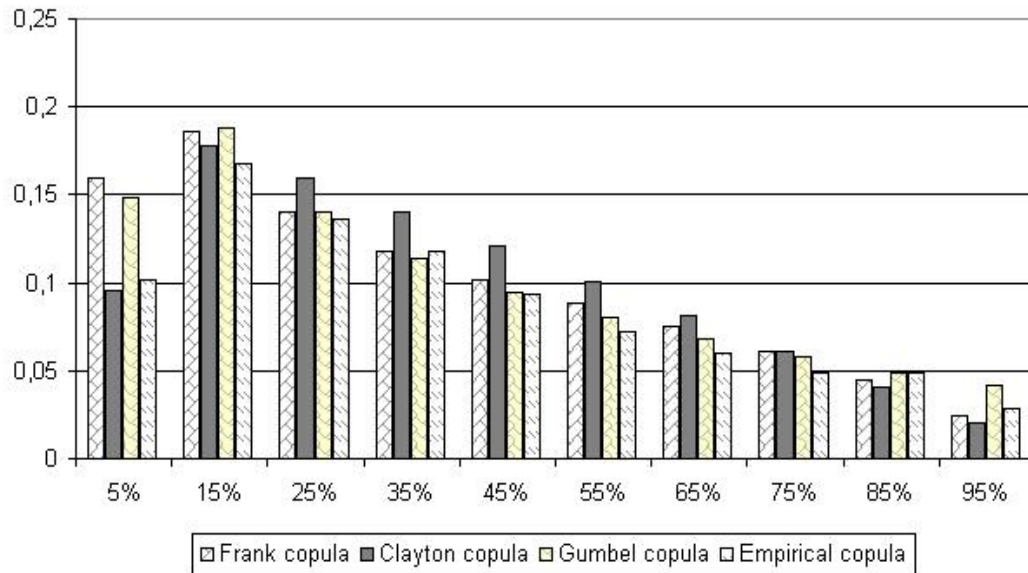
Figure 12: Comparison of the distribution (ie the function K) of the copula function for the HSBC-RBS couple

distribution. A similar result has been described by Gatfaoui [10]. In an article studying the correlation between index returns and credit spreads, she concluded that Frank copula was the best copula to describe the dependence structure between index returns and credit spreads. The interest of this conclusion is that we can now use the Frank copula function to model the dependence structure between stock returns, and as a proxy between default probabilities. Then, we can conclude that the market price we should obtain for a basket CDS is closer to the price obtained when we model the dependence structure of the portfolio with a Frank copula.
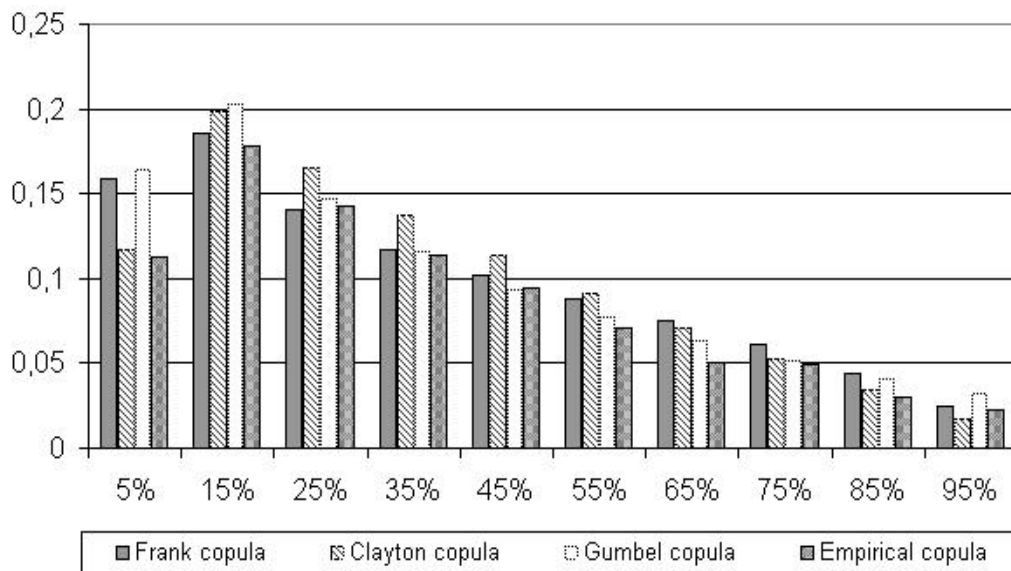
Figure 13: Comparison of the distribution (ie the function K) of the copula function for the HSBC-BP couple

Conclusion

The description of dependence is paramount in finance. Indeed, dependence is often described as a mere number which is the correlation coefficient and seldom described more completely like the structure which can be achieved using copula functions. Moreover, the normal multivariate distribution is still widely used in finance, whereas it does not describe accurately the behavior of portfolios.

In this thesis, we have introduced some tools to understand the basic concepts of the copula function theory which makes it possible to model this dependence structure precisely. Indeed, we have seen that the family of the copula functions is a very wide family, where each copula describes a different dependence structure. We have particularly focused our attention on the Archimedean copula functions because this family of copula functions is easily tractable and has many interesting properties. Moreover, we have seen that the semi-parametric estimation is a very powerful and tractable tool to estimate copula functions and compare the different dependence structures they model. Finally, we explained the methods to realize the choice of the best copula function given a set of data, using the empirical copula functions.

In order to understand the application of copula functions in the pricing of credit derivatives, we applied most of the results demonstrated in this thesis in order to realize the pricing and the study of a basket CDS. Throughout those applications, we studied the impact of the different parameters of the portfolio. We particularly studied the impact of the dependence structure.

References

[1] S. Avouyi-Dovi and D. Neto. Les fonctions copules en finance. *Banque et Marchés*, (68):44–57, 2004.

[2] C. Bluhm, L. Overbeck, and C. Wagnen. *An introduction to credit risk modelling.* Chapman and Hall, 2003.

[3] E. Bouyé, V. Durrlemann, A. Nikeghbali, G. Riboulet, and T. Roncalli. Copulas for finance: A reading guide and some applications. *Working Paper*, July 2000.

[4] D. Cadoux and J-M. Loizeau. Copules et dépendances: application pratique à la détermination du besoin en fonds propres d'un assureur non vie. *Working Paper*, 2003.

[5] P. Deheuvels. A non parametric test for independance. *Publications institutionnells de Statistiques Université de Paris*, 2:29–50, 1981.

[6] P. Embrechts, A. McNeil, and D. Straumann. Correlation and dependency in risk management: properties and pitfalls. *Working Paper*, November 1998.

[7] J-D. Fermanian and O. Scaillet. Nonparametric estimation of copulas for time series. *Working Paper*, 2003.

[8] J-D. Fermanian and O. Scaillet. Some stastical pitfalls in copula modeling for financial applications. *Working Paper*, 2004.

[9] M-J. Frank. On the simulation associativity of f(x,y) and x+y-f(x,y). *Aequationes Mathematicae*, 19:194–226, 1979.

[10] H. Gatfaoui. How does systematic risk impact us credit spreads? a copula study. *Working Paper*, 2003.

[11] C. Genest and J. MacKay. Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *The Canadian Journal of Statistics*, 14:154–159, 1986.

[12] C. Genest and L. Rivest. Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, 88:1034–1043, 1993.

[13] Younés Hillali. Analyse et modélisation des données probabilistiques: capacités et lois multidimensionnelles. *Université Paris IX Dauphine*, 1998.

[14] L. Hu. Dependence patterns across financial markets: a mixed copula approach. *Working Paper*, June 2004.

[15] J. Hull and A. White. Valuation of a CDO and an n-th to default CDS without Monte-Carlo simulation. *The Journal of Derivatives*, 12(2):8–23.

[16] H. Joe and J. Xu. The estimation method of inference functions for margins for multivariate models. *Working Paper*, 1996.

[17] J.-F. Jouanin, G. Rapuch, G. Riboulet, and T. Roncalli. Modelling dependences for credit derivatives with copulas. *Working Paper*, August 2001.

[18] D.X. Li. On default correlation: a copula function approach. *Journal of Fixed Income*, 9(4):43–54.

[19] Lee McGinty, Eric Beinstein, Rishad Ahluwalia, and Martin Watts. Credit correlation: A guide. *JPMorgan Credit Derivatives Strategies*, March 2004.

[20] Robert C. Merton. On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, 29:449–470.

[21] R.B. Nelsen. *An introduction to copulas.* Springer-Verlag New-York, 1998.

[22] T. Roncalli. *Gestion des risques multiples ou copules et aspects multidimensionnels du risque.* Cours ENSAI de 3eme année, 2002.

[23] A. Sklar. Fonctions de repartition a n dimensions et leurs marges. *Publication Institutionnelles Statistiques Université Paris*, 8:229–231, 1959.

[24] A. Sklar. Random variables, joint distribution functions and copulas. *Kybernetika*, 9:449–460, 1973.