MODELING THE FINANCIAL MARKET USING COPULA

A Thesis

Presented to

The Graduate Faculty of The University of Akron

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

Michael Gyamfi

May, 2017
This project is to track the differences and the movements between the Actual and theoretical future prices using Copula. SPX and 10-year treasury bond yield rate was downloaded from Yahoo! website and SPX future prices were downloaded from Moore Research Center website and their observations from January 2, 2001 to May 27, 2016 were used for this analysis. Log-returns of the future prices were taken to model and analyze the direct movements of the future prices. The distributions of the marginals and the best family of copula was selected and simulated. We compared the copula method to the classical method after 2000 simulation.

A high level of mis-pricing in the future price which corresponds to the period 2008-2009 was observed. This observed mis-pricing could be as a result of relative over-reaction of the Financial market compared to future market. Inverse relationship between the performance of SPX and the volatility of future prices was observed. Standardized Student’s t-distribution was concluded to be the marginal distribution using the maximum likelihood method to estimate their distribution parameters. Student t-Copula was concluded to be the best family of copula to measure the dependence. In further studies, modeling the risk associated with futures stock price and pricing with copula based simulation will be a major red flag to be addressed.
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>vi</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
</tbody>
</table>

### CHAPTER

#### I. INTRODUCTION

1.1 Shortcomings of Pearson Correlation Coefficient .................................. 2

1.2 Tail Dependence in Finance ................................................................. 4

1.3 S&P 500 and Volatility Index ............................................................. 5

#### II. LITERATURE REVIEW

2.1 Cathy, Dinghai and Tony (2008) ......................................................... 7

2.2 Yiguo Sun and Ximing Wu (2006) ......................................................... 8

2.3 Fountain, Herman and Rustvold (2008) ................................................ 9

2.4 Sriboonchitta, Nguyen et al. (2013) ..................................................... 9

2.5 Christoffersen, Jacobs, Jin and Langlois (2014) .................................... 11

#### III. COPULA

3.1 DEFINITION OF COPULA ................................................................. 13

3.2 TYPES OF COPULAS ................................................................. 16

3.3 MEASURE OF DEPENDENCE ......................................................... 20
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>30</td>
</tr>
<tr>
<td>4.2</td>
<td>30</td>
</tr>
</tbody>
</table>

4.1 Parameter estimates for Standardized Student’s T distribution used to model $X_i$ and $Y_i$.

4.2 Estimated Parameters Using MLE.
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Actual Future vs Theoretical non-arbitrage Future price</td>
<td>24</td>
</tr>
<tr>
<td>4.2</td>
<td>Performance of SPX in log scale.</td>
<td>25</td>
</tr>
<tr>
<td>4.3</td>
<td>Performance of VIX.</td>
<td>26</td>
</tr>
<tr>
<td>4.4</td>
<td>Difference between Theoretical and Actual Future Price in percentage</td>
<td>27</td>
</tr>
<tr>
<td>4.5</td>
<td>Scatter plot of log-returns of theoretical future price and actual future price.</td>
<td>28</td>
</tr>
<tr>
<td>4.6</td>
<td>Empirical distribution and fitted standardized Student’s T distribution of $X_i$ and $Y_i$.</td>
<td>29</td>
</tr>
<tr>
<td>4.7</td>
<td>CDF of theoretical Vs CDF of Actual Future price</td>
<td>31</td>
</tr>
<tr>
<td>4.8</td>
<td>Observed vs Simulated using t-copula (above) and bivariate t-distribution. There are 3852 observations, and 2000 simulated points in each plot. Both models have similar standard deviations, correlation, and Kendall’s tau.</td>
<td>33</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

The financial crisis which began in the second half of 2007 has been characterized by the transmission of shocks across various financial markets and inter-linkages between and co-movement of differing assets classes like stocks, bonds and commodities [1]. These dislocations were initially caused by a shock in the U.S housing market triggered by cheap credits and subsequent reversal of interest rate and falling house prices [1]. This shocks spread across the asset classes leading to disruptions of the interbank money, credit market, foreign exchange and other financial markets [1]. Financial markets tend to exhibit lower tail dependence and major stock returns in normal times had lower correlation before the crisis. In a crisis, financial correlation typically increase because the tail dependence and extreme values are not accounted [2]. Tail dependence is the measure of the extreme co-movements among financial variables. Correlation measures the degree to which prices of assets move together. Detailed definition of tail dependence is given in the next chapter. However, during the crisis most of the asset values dropped in conjunction with the equity values and the correlation among these assets rose significantly. Many investors as a result realized that portfolios which they believed to be well diversified based on historical data were effectively not diversified at all. The traditional measure of dependence such as Pearson
Correlation Coefficient failed to model correctly the variables in the financial market. This incident led to the popularity of the use of Copula in the financial market. A copula is a function that links or joins univariate marginals to their multivariate distribution [3].

In this paper, we develop Copula models to directly measure the interrelationship between the stock returns and its future price. Stock returns are the income or capital gains or loss on an investment in the stock market [4]. Futures are financial contracts obligating the buyer and seller to buy and sell an asset at a fixed future date and price [4]. We will examine the structure of the individual variable data sets and evaluate the association between the Standard and Poors (S&P) 500 Index (SPX) and its future contracts using 10 years treasury bond data set from Yahoo website and SPX future prices data from Moore Research Center website.

The remainder of this paper is organized as follows. The rest of this chapter lists further motivations for this project. Chapter 2 presents a survey of similar topics in the literature. Chapter 3 introduces the Copula function, and finally the data analysis is presented in Chapter 4.

1.1 Shortcomings of Pearson Correlation Coefficient

Pearsons or Gaussian correlation has been at the center for modeling and measuring dependencies historically. Pearsons Correlation coefficient between two random
variables $X$ and $Y$ with finite variance is defined as:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{V}(X)\text{V}(Y)}}. \quad (1.1)$$

Where $\text{Cov}(X,Y) = E(X,Y) - E(X)E(Y)$ represents the covariance between the random variables, and $\text{V}(X)$and $\text{V}(Y)$ are the variances of the random variables $X$ and $Y$ respectively. Pearson’s correlation coefficient satisfies: $-1 \leq \rho(x, y) \leq 1$, and $\rho(x, y)$ will take values of $\pm 1$ only with a perfect linear dependence. Pearson’s correlation coefficient represents the strength of a linear relationship between two random variables. Lack of correlation is only a necessary condition for independence, and only in the case of elliptically distribution, Pearson coefficient can be reliably used [5]. Elliptical distribution is an extension of a normal distribution. The joint distribution of an elliptical distribution forms an ellipse or ellipsoid, in a two or three dimensional case [?]. The linear correlation coefficient is not invariant with respect to non-linear (e.g. log) transformation, and if non-linear (e.g. quadratic or cubic) dependence between variables can not be represented. This fact is demonstrated by this example.

Suppose a standard normal random variable $X \sim N(0,1)$ with quadratic dependence $Y = X^2$ is investigated. Despite this direct relationship between $X$ and $Y$, calculation of the correlation coefficient shows that

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0, \quad (1.2)$$

because $E(X) = 0$, $E(Y) = E(X^2) = V(X) = 1$, and $E(XY) = E(X^3) = 0$. This means that $\rho(X, Y) = 0$ and this is one example where Pearson’s or Gaussian corre-
lation may exhibit no correlation if there is non-linear relationship between random variables.

In past times, the Pearsons correlation was commonly used as a measure for price of the home credit products and since it did not account for tail dependence or non-linear dependence, it under-estimated aggregated risk of potential losses. Random variables representing home-owner’s payments exhibited low correlation in the normal times, but proved to be highly correlated in an adverse economic climate. In other words, tail dependence of the random variables was neglected, and over-simplifying the correlation coefficient was being used as the basis for the risk measure.

1.2 Tail Dependence in Finance

The tail of a distribution is the part of a distribution that is very far away from the mean or the center of the distribution. Tail dependence between two random variables is observed when the correlation between two variables increases as you get further to the tail of the distributions. The concept of tail dependence relates to amount of dependence in the upper right quadrant tail or lower left quadrant tail of a bivariate distribution. Random variables that appear to exhibit no correlation when $X$ and $Y$ are close to $E(X)$ and $E(Y)$ respectively, can show great dependence when $X$ and $Y$ are both far away from their respective means. It is a stylized fact of stock returns exhibit tail dependence [6].

This concept is relevant for the study of dependence between extreme values. For example, take the re-insurer of credit risk for instance. When the overall losses
are normal, whether re-insurer A or re-insurer B will default on their payments to an insurer may look uncorrelated, or very weakly corrected. Now when re-insurance is more likely to be claimed when a string of catastrophes happen (e.g. Hurricanes Rita, Wilma, and mass car destruction etc). In such a scenario, the entire market will be hit one after the other with huge request for payments, and simultaneous demands of their insured.

1.3 S&P 500 and Volatility Index

The Chicago Board of Exchange (CBOE) Volatility Index (VIX) has been widely accepted as a measure of "fear" present in the US stock market. Designed in 1993, it was first calculated from the implied volatility of out-of-the-money S&P 100 index options contract prices and later it was redefined to be based using options contract prices of the S&P 500 index.

In general, VIX tend to increases when the market falls, and decreases when the market goes up. For example, during the 2008 financial crisis, VIX recorded highest value ever to date of 59.89 (Oct 1, 2008). Five months earlier it was 17.83 (May 1, 2008), which is much smaller but still considered to be relatively high and "alarming" compared to values recorded in more tranquil times such as 10.42 (Jan 1, 2007) and 11.99 (Jan 1, 2017).

Capturing the exact relationship between S&P 500 index and VIX index are of great interest, as VIX index can be an ideal risk hedging vehicle in the time of crisis. There is much literature investigating the dependence between daily movements in
S&P 500 index and that of VIX index. The next chapter lists synopses of some of them.

Instead of repeating the modeling dependency between the S&P 500 index and VIX index, this project aims to model daily movement of S&P 500, and daily movement in settlement price of the closest monthly S&P 500 future contract. Similar to the S&P 500 option contracts prices that are basis for the VIX index, S&P 500 future prices also reflect anticipation and expectation for the near-term future movement of S&P 500 index. Studying the relationship between movements of S&P 500 index and its future values is not only interesting for its own right, it is hoped to shed light on better modeling of dependency between S&P 500 index and VIX index.
CHAPTER II
LITERATURE REVIEW

This chapter contains review of papers that investigates the relationship of stock price volatility to stock price returns using copula function. All of them reports the general theme of the high negative correlation between the S&P 500 index and the VIX index. Copula functions are used in most of the literature to model this relationship between S&P 500 index and VIX index.

2.1 Cathy, Dinghai and Tony (2008)

[7] This paper developed copula models to directly measure the inter-relationship between the stock returns and its individual future volatility. They used empirical evidence which suggests that the leverage effect tends to occur as a downside effect. Leverage effect shows the relationship between stock returns and both implied and realized volatility [8] This implies that only a large negative return is followed by an increase in its volatility. Traditional measure of correlation cannot capture this form of leverage effect. Copula models, on the other hand, can directly measure this effect by measuring the tail dependence of the return and its volatility distribution. [7] To model the leverage effect, they applied various copula models. They started with a mixture of the Clayton and survival Clayton copula. Survival Clayton copula
happens when the estimation results indicate a zero weight on the Clayton copula. Thus, they focused on the copulas naturally belonging to the Gumbel copula and survival Clayton copula and they used them in the analysis of paper. In addition, they allowed for time varying in the tail dependence of the return distribution. In particular, following [9], they propose the following ARMA (Autoregression Moving Average) type process for the innovation of the tail dependence.

2.2 Yiguo Sun and Ximing Wu (2006)

[10] The authors studied the relationship between the S&P 500 index returns and the returns of VIX via nonparametric copula method. The paper further proposed a conditional dependence index to investigate how the dependence between the two return series varies across different segments of market return distribution. Their observed findings are as follows: a) the two series exhibit strong, negative, extreme tail dependence; b) the negative dependence is stronger in extreme bearish markets than in extreme bullish markets; c) the dependence gradually weakens as the market return moves toward the center of its distribution, or in quiet markets. The unique dependence structure supports the VIX as a great measure of markets’ mood in general. Finally, they proposed a stylized model of the returns of the VIX based on the S&P 500 index returns that incorporates the asymmetry and tail dependence between the two series.
2.3 Fountain, Herman and Rustvold (2008)

[11]

The authors showed the dependence orderings of the SPX and VIX in two different ways: first, they provided alternative measures of the dependence between SPX and VIX, which, although varying in value, agreed in sign with the usual coefficients of correlation. Second, they provided a method of gauging the amount of tail dependence between SPX and VIX, through comparisons with the Gumbel family of copulas. They hoped to do a further study to determine which of the meta-copulas provides the most reasonable probability measure for the comparison of Kendall random variables, and it may be that certain meta-copulas are optimal in specific settings. [12] defined an ordering of two-dimensional random vectors based on their Kendall distributions. Based on that ordering, they defined the notion of positive Kendall dependence. However, their ordering required the stochastic domination of one Kendall random variable over another. Fountain and Rustvold's method removed this requirement, thus provided a more general means of comparison.

2.4 Sriboonchitta, Nguyen et al. (2013)

[13] The paper modelled volatility and dependence of agricultural price and production indices of Thailand, using static versus time-varying copulas. The study describes the behaviours of conditional volatilities of growth rates for agricultural prices and production. These estimated conditional volatilities can help policy mak-
ers understand the behaviours of volatilities and subsequently make possible policy adjustments to prevent undesirable effects caused by the change in price and production. Their results showed that the estimated dependence parameter was relatively low, approximately -1. They found that the appropriate marginal density for growth rates of agricultural production and price indices were the skewed t-distributions with the means, volatilities, skewness and degrees of freedom. The time-varying rotated Joe copula was the best among several copula candidates.

2.4.1 Sukcharoen, Zohrabyan, Leatham, and Wu (2009)

[14] The study applied the copula approach to model the interdependence of oil prices and stock market indices. Their study considered the following nine different parametric copula functions from various copula families and classes: Gaussian (Normal), Clayton, Rotated Clayton, Frank, Plackett, Gumbel, Rotated Gumbel, Students t, and Symmetrized Joe-Clayton (SJC). The most appropriate copula function for each bivariate model was selected based on the log-likelihood functions and two information metrics. They address specifically the following questions: do oil prices and stock market indices move together? Is there any asymmetry in the relationship? Does the dependence (if any) increase during extreme events? Does the dependence change for pre and post Euro periods? Is there a specific dependence pattern for developed and developing countries and for oil producing and consuming countries?
2.5 Christoffersen, Jacobs, Jin and Langlois (2014)

[15] The authors characterized dependence and tail dependence in corporate credit using a new class of dynamic copula models to capture dynamic dependence and asymmetry in large samples of firms. They noted the importance of the differences between the dependence dynamics for credit spreads and equity returns. They found that relationships are highly time-varying, which increased significantly in the financial crisis. They concluded that the CDS volatility, correlation and tail dependence measures that have constructed using the dynamic copula model are important determinants of credit spreads over time.

Financial risk is mostly composed of rare or extreme events which results in high risk and lies in the tail of return distribution. In option pricing, rare or extreme events results in volatility skew patterns [16]. [17] [18], and [19] used Ordinary Least Square in their study of asymmetric return-volatility relationship across implied volatility change distribution.

From the literature cited, there is a central conclusion that the S&P 500 and VIX indices are negatively related. Also, their correlation or relationship exhibits extreme values and tail dependence which cannot be modeled by Pearson correlation. The next chapter introduces Copula and other dependence measure that can be used to better model the market structure.
The word copula is from the Latin word copulare, which means to join together. Theoretically, complete details of dependency between two random variables are captured in the joint cumulative distribution function (CDF) or joint probability density function (pdf). However, estimation of joint CDF or pdf suffers ”curse of dimensionality” which means it requires large number of observations even when parametric family of the random variables are assumed [20]. This difficulty has traditionally forced researchers to use parametric form of the bivariate or multivariate distribution. One drawback of such approach is that it requires $X$ and $Y$ to have same marginal distribution. Another drawback is that there are limited number of bivariate parametric family of distributions for researchers to adopt to the data at hand.

Copula is a function that ties two marginal cumulative distribution functions together, to represent a dependency structure. In the copula model, marginal distribution of each random variable can be chosen separately, and while one particular copula is restricted to a certain form of dependency, there are many types of copula functions available to pick and choose from, to suit the dependency present in your data set.

For sake of simplicity of our discussion, we restrict our attention to the bi-
variate case. However the definitions and methods presented below extend naturally to higher dimensional data.

3.1 DEFINITION OF COPULA

A copula is a function that joins multivariate joint CDF to their marginal univariate CDF. In the representation theorem, [21] states that the joint CDF $H(x, y)$ of any pair $(X, Y)$ of continuous random variables may be written in the form

$$H(x, y) = C\{F(x), G(y)\}, \quad x, y \in \mathbb{R} \quad (3.1)$$

where $F(x)$ and $G(y)$ are marginal CDFs of random variable $X$ and $Y$ respectively. $C : [0, 1]^2 \rightarrow [0, 1]$ is a copula function. As Sklar has shown in the representation theorem in 1959 [21], when $H(x, y)$ is known, $C, F, G$ are uniquely determined.

As one can see in above definition, this allows different modeling approach to given data. First, we model distribution of $X$ and $Y$ separately. Once we find reasonable estimate for $F$ and $G$, the dependency between the two can be modeled via copula function $C$.

3.1.1 Invariance of Copula Under Transformation

There is another desirable property of copula, namely the invariance under monotone increasing transformation of marginals.

Suppose random variables $X$ and $Y$ are dependent upon the relationship

$H(x, y) = C\{F(x), G(y)\}$. Suppose further, that $\phi$ and $\psi$ are monotone increasing
transformations. Then the joint CDF of the transformed random variables $\phi(X)$ and $\psi(Y)$, can be written as

$$H^*(u, v) = C^*[F^*(u), G^*(v)]$$

(3.2)

where $F^*$ and $G^*$ are CDFs of transformed random variables. Observe that

$$F^*(x) = P(\phi(X) \leq x) = P(X \leq \phi^{-1}(x)) = F(\phi^{-1}(x)),$$

(3.3)

and similarly $G^*(y) = G(\psi^{-1}(y))$. That means we can write the joint CDF of transformed random variables as

$$H^*(u, v) = P(\phi(X) \leq u, \psi(Y) \leq v)$$

(3.4)

$$= P(X \leq \phi^{-1}(u), Y \leq \psi^{-1}(v))$$

(3.5)

$$= H(\phi^{-1}(u), \psi^{-1}(v))$$

(3.6)

$$= C[F(\phi^{-1}(u)), G(\psi^{-1}(v))]$$

(3.7)

$$= C[F^*(u), G^*(v)].$$

(3.8)

proving that $C^* = C$. So the copula that binds $F$ and $G$ to form $H$ is the same copula function that binds $F^*$ and $G^*$ to form $H^*$.

This is a desirable property for dependence modeling. Pearson Correlation is invariant under linear transformation, but not under nonlinear transformation.

Given bivariate data $(X_1, Y_1), \ldots, (X_n, Y_n)$, scatter plot is often shown to visualize the dependency structure. When the dependency is modeled with copula function, because of this invariance property, copula function can be visualized by

14
plotting scatter plot of

\[ \left( \hat{F}(X_1), \hat{G}(Y_1) \right), \ldots, \left( \hat{F}(X_n), \hat{G}(Y_n) \right). \]  \tag{3.9}

Assuming \( \hat{F} \) and \( \hat{G} \) are close to \( F \) and \( G \), transformed random variables \( F(X_i) \) and \( G(Y_i) \) have uniform distributions, whose CDF is \( F^*(u) = u \) and \( G^*(v) = v \). Therefore, transformed scatter shows

\[ H^*(u, v) = C[F^*(u), G^*(v)] = C[u, v]. \]  \tag{3.10}

Which is the copula function itself.

### 3.1.2 Frechet-Hoeffding Bounds Inequality

Let \( C \) be a copula. Then for every \( (u, v) \) in \( \text{dom}C \),

\[ \max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v) \]  \tag{3.11}

The bounds in the above inequality are themselves copulas and are commonly denoted by \( M(u, v) = \min(u, v) \) and \( W(u, v) = \max(u + v - 1, 0) \). In order for random variable \( Y \) to be a deterministic function of \( X \), \( C \) must be with \( M \) or \( W \) [22]. This bounds are analogous to bounds for Pearson Coefficient \( \pm 1 \), which only happens when \( Y \) is linear deterministic function of \( X \) [22].

Another important special case copula is the product copula \( \pi(u, v) = uv \).

Random variables \( X \) and \( Y \) are independent if and only if \( C = \pi \).
3.2 TYPES OF COPULAS

There have been many different classes of copulas developed over the years. Here, two important ones are defined: Elliptical and Archimedean Copula. Archimedean copula have the advantage of having only one parameter. Also, Archimedean Copula can easily be constructed, have many varieties available and they possess nice and flexible properties compared to the elliptical copulas.

3.2.1 Elliptical Copulas

Elliptical copulas are simply the copulas of elliptical distributions. Elliptical distributions provides a great source of multivariate distribution which shares many properties of multivariate normal distribution and enables modeling of multivariate extremes and other forms of non-normal dependence [23]. From [6] an elliptical distribution is any member of a broad family of probability distribution that generalize the multivariate normal distribution. Below, we present two examples of Elliptical Copula; the Gaussian Copula and the Student’s t-Copula.

Gaussian Copula

The Gaussian copula is defined as:

\[ C(u, v; \rho) = \Phi_{2,\rho}(\Phi^{-1}(u), \Phi^{-1}(v)) \]  

(3.12)
where $\Phi_{2,\rho}$ is the bivariate distribution with the correlation $\rho$, and $\Phi^{-1}$ is the inverse of marginal standard normal CDF (i.e. quantile function). Note that $\rho$ is simply the usual linear correlation coefficient of the corresponding bivariate normal distribution.

Student’s t-Copula

Similar to the Gaussian Copula, t-copula models the dependence structure of multivariate t-distributions, which is defined similarly as:

$$C_{v,R}^t(u) = \Phi_{v,R}^t\left(\Phi_{v}^{-1}(u_1),...,\Phi_{v}^{-1}(u_n)\right)$$  \hspace{1cm} (3.13)

The parameters for the student t-copula are the correlation $\rho$ and degrees of freedom $nu$. Student t-copula shows the symmetrical dependence, but is higher than those in Gaussian copula.

3.2.2 Archimedean Copulas

Archimedean copulas are well-known compared to elliptical mostly because they allow modeling dependence in arbitrarily high dimensions with only one parameter that controls the degree of dependence. To define Archimedean copula, the notion of pseudo-inverse must be introduced. Let $\varphi$ be a continuous, strictly decreasing function from $[0, 1] \rightarrow [0, \infty]$ such that $\varphi(1) = 0$. $\varphi^{-1}$ is called pseudo-inverse of $\varphi$ and given by:

$$\varphi^{-1} = \begin{cases} 
0 & \text{if } \varphi(0) \leq t \leq \infty \\
\varphi^{-1}(t) & \text{if } 0 \leq t \leq \varphi(0)
\end{cases} \quad \text{for all } t \in [0, \infty]. \hspace{1cm} (3.14)$$
With this pseudo-inverse, Archimedean copula can be defined.

From [24] and [22], a 2-dimensional copula $C$ is called an Archimedean Copula if and only if there exists a continuous, strictly decreasing, convex function $\varphi$ (called Generator function) which maps from $[0, 1]$ to $[0, \infty]$ such that $\varphi(1) = 0$;

$$C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2)) \quad (3.15)$$

For this paper, we present the three most popular Archimedean copulas which include Clayton, Gumbel and Frank Copulas.

Clayton Copula

The Clayton copula is mostly used to study correlation risks because of its ability to capture lower tail dependence. The closed 2-dimensional form of the bivariate Clayton copula is defined as:

$$C^c(u, v; \theta) = \max\left(u^{-\theta} + v^{-\theta} - 1, 0\right)^{-1/\theta} \quad (3.16)$$

Where $\theta$ is the copula parameter strictly within the interval $(0, \infty)$. If $\theta = 0$, then the marginal distributions become independent. Clayton copula approximates the Frechet-Hoeffding upper bound when $\theta \to \infty$. From the restriction on the dependence parameter, Frechet-Hoeffding lower bound cannot be found by Clayton copula.
Gumbel Copula

The Gumbel Copula is used to capture strong upper tail dependence and weak lower tail dependence. A situation where the outcomes are expected to be strongly correlated at high values but less correlated at low values, then the Gumbel copula is an appropriate choice. The Gumbel copula is given by:

$$C^G(u, v; \theta) = \exp \left[ - ( - \log(u)^\theta) + ( - \log(v)^\theta)^{\frac{1}{\theta}} \right],$$

where $\theta$ is the copula parameter within the interval $[1, \infty)$. When $\theta$ approaches 1, the marginals become independent and when $\theta$ goes to infinity the Gumbel copula approaches the Frechet-Hoeffding upper bound. Similar to the Clayton copula, the Gumbel copula represents only the case of independence and positive dependence.

Frank Copula

The Frank copula is defined as:

$$C^F(u, v; \theta) = -\frac{1}{\theta} \log \left[ 1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{(e^{-\theta} - 1)} \right]$$

Here $\theta$ is the copula parameter that may take any real value. Unlike the Clayton and the Gumbel copula, the Frank copula allows the maximum range of dependence. That is, the Frank copula dependence parameter permits the approximation of the Frechet-Hoeffding upper and lower bounds and allows modeling positive as negative dependence. The Frechet-Hoeffding upper and lower bounds are attained when $\theta$ approaches $\pm \infty$ and the independence case is attained when $\theta$ approaches 0. Therefore,
the Frank Copula has neither lower nor upper tail dependence. The Frank copula is often used to model data set characterized by weak or no tail dependence.

3.3 MEASURE OF DEPENDENCE

Below, we define two alternative nonparametric or nonlinear measures of dependence between the two variables, namely the Spearmans $\rho$ and Kendalls tau $\tau$. Unlike the Pearsons correlation coefficient, these rank correlations do not require a linear relationship between the variables. Both Spearmans rho($\rho$) and Kendalls tau ($\tau$) correlation coefficients measure a form of dependence known as concordance.

Concordance

Informally, a pair of random variables are concordant if large values of one tend to be associated with large values of the other and small values of one with small values of the other. More precisely, let $(x_i, y_i)$ and $(x_j, y_j)$ denotes 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 $x_i > x_j$ 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 $y_i < y_j$. The alternative 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 $(x_i - x_j)(y_i - y_j) < 0$. 

20
Kendalls tau ($\tau$)

The Kendalls tau measures the difference between the probability of concordance and probability of discordance. [25]. Let $(X_1, Y_1)$ and $(X_2, Y_2)$ be independent and identically distributed random vectors, each with joint distribution function $H$. Then the population version of Kendall’s tau is defined as:

$$
\tau_{X,Y} = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] = \frac{c - d}{nC_2} \quad (3.19)
$$

Where $c$ denotes the number of concordant pairs and $d$ represents the number of discordant pairs. There are $nC_2$ distinct pairs of $(X_1, Y_1)$ and $(X_2, Y_2)$ observations in the population and each pair either a concordant or discordant.

Spearmans rho ($\rho$)

Let $(X_1, Y_1)$, $(X_2, Y_2)$ and $(X_3, Y_3)$ be three independent random vectors with common joint distribution function $H$, whose margins are $F$ and $G$ and Copula $C$. The population version $\rho_{X,Y}$ of 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 $(X_2, Y_3)$. Mathematically,

$$
\rho_{X,Y} = 3 \left( P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0] \right). \quad (3.20)
$$
CHAPTER IV
DATA ANALYSIS AND MODELING

This chapter analyzes daily movement in S&P 500 index (SPX) and its nearest future contract. SPX and 10-year treasury bond yield rate was downloaded from Yahoo! website. SPX future prices were downloaded from Moore Research Center website. We analyze observations from January 2, 2001 to May 27, 2016.

The no-arbitrage principle in Finance states that any future contract that expires in $D$ days of any stock with current price $S_0$ must be priced at

$$ F_T = S_0 \left(1 + R\right)^{\frac{D}{365}}, \quad (4.1) $$

Where $R$ is the effective annual risk-free interest rate. Otherwise, there will be an arbitrage, which means that there will be a profitable transaction that is risk-free. We will use 10-year treasury bond rate as risk-free interest rate.

Letting $i$ be the index for $i$-th day, we denote our theoretical future price sequence as

$$ F^T_i = S_i \left(1 + R_i\right)^{\frac{D_i}{365}}, \quad (4.2) $$

where $S_i$ denotes the SPX index closing price for $i$-th day, $R_i$ denotes the risk-free rate, and $D_i$ is the number of days remaining until the nearest future contract expiration date. Our interest is to track the differences and movements between the
where \( F_i \) is the actual, observed daily closing price for the nearest expiration SPX future contract. Figure 4.1 shows the scatter plot of \((F_i, F^T_i)\). Market does seem to care about the arbitrage, and it seems as if \( F_i = F^T_i \). However, if you zoom in as shown on the right hand side of Figure 4.1, it is not quite on the line. From this plot it seems that on some days, there are about 1% difference between the actual future contract price and theoretical future price. It is worth noting that the theoretical price is almost always above the actual price. In other words, SPX futures tend to be "under-priced".

4.1 Theoretical vs Actual Future Prices

Figure 4.2 shows performance of SPX from January 2001 to May 2016. We can observe so-called "tech bubble burst" downward trend during the year 2003 and 2004. There is also a largest decline in the stock market during the 2008 and early 2009 subprime financial crisis. Figure 4.3 below plots reaction in VIX index. The plot shows that VIX spikes up at the same time as SPX goes down.

Figure 4.4 shows the plot of daily difference in actual and theoretical future price in percentage,

\[
\Delta_i = \frac{(F_i - F^T_i)}{F_i}.
\]
As well as the periodical nature of mis-pricing, the plot shows that the high level of mis-pricing in future price corresponds to the period of 2008-2009 Financial Crisis. These mis-pricing can be explained as relative over-reaction of the financial market compared to future market, in the event of the crisis. The periodic nature of the mis-pricing is due to the expiration date. As the expiration of the future contract comes near, the difference between actual and theoretical prices becomes small. On the date of the expiration, the two must be the same.

![Graph showing Actual Future vs Theoretical non-arbitrage Future price](image)

Figure 4.1: Actual Future vs Theoretical non-arbitrage Future price
4.2 Log-return of the Future Prices

We now turn our attention to how the two future prices, actual and theoretical, move from day to day. Since the theoretical price is directly related to the movement of SPX, studying the relationship between the two movements is one way of studying the "reaction" of the future market to the index movements. We first plot, in Figure 4.5, log-returns of $F_i$ and $F_i^T$,

$$\left( \log\left(\frac{F_i}{F_{i-1}}\right), \log\left(\frac{F_i^T}{F_{i-1}^T}\right) \right)$$

Pearson’s Correlation coefficient for this relation was calculated as 0.9777, and
Figure 4.3: Performance of VIX.

Kendall Tau was calculated as 0.8667.
Figure 4.4: Difference between Theoretical and Actual Future Price in percentage
Figure 4.5: Scatter plot of log-returns of theoretical future price and actual future price.
4.3 Modeling Marginal Distributions

Before we apply the copula function to model relationship shown in Figure 4.5, we must model the marginal distribution of log-return of actual future price and theoretical future price independently. For the sake of the brevity, we denote

\[ X_i = \log\left(\frac{F_i}{F_{i-1}}\right), \quad Y_i = \log\left(\frac{F_{T_i}}{F_{T_{i-1}}}\right). \quad (4.6) \]

For the marginal distribution, standardized Student’s t-distribution was chosen. Distribution parameter was estimated by maximum likelihood method. Figure 4.6 shows empirical distribution of \( X_i \) and \( Y_i \) in black, and fitted distribution function in red. The plot shows good agreement between the two. Parameter estimates for the two distributions are listed in Table 4.1.

![Empirical distribution and fitted standardized Student’s T distribution](image)

Figure 4.6: Empirical distribution and fitted standardized Student’s T distribution of \( X_i \) and \( Y_i \).
4.4 Modeling with T-Copula

Figure 4.7 below shows the scatter plot of \((\hat{F}(X_i), \hat{F}(Y_i))\), where \(\hat{F}\) and \(\hat{F}\) are the ECDFs. From this scatter plot, we can observe high level of dependence at its tails. This plot shows a symmetric tail dependence and we concluded to use the student t-copula in modelling distributions with symmetric tail dependence. The t-copula is usually good for modelling phenomena where there is high correlation in the extreme values.

We continue to estimate the parameters of the t-copula using the maximum likelihood estimation (MLE). The MLE approach estimates the t-copula parameter and the parameters of the marginal distribution simultaneously. The estimated parameters are: where \(\rho\) is the correlation between the marginals and \(df\) is the degree.

Table 4.2: Estimated Parameters Using MLE

<table>
<thead>
<tr>
<th></th>
<th>(\rho)</th>
<th>(df)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.9786</td>
<td>1.9387</td>
</tr>
</tbody>
</table>
of freedom using the t-copula. Next section will verify how well the t-copula model captures the dependency.

4.5 Simulating from t-Copula and Comparison to Bivariate t-distribution

In order to show that t-Copula fitted captures the dependency exhibited by X and Y in Figure 4.7, we simulate 2000 random sample from the t-Copula distribution with parameter in Table 4.2 with marginal standardized t-distribution with parameters in Table 4.1. Result is shown at the top of Figure 4.8.
In order to make a comparison with classical method, we have also fitted $X$ and $Y$ with bivariate t-distributions. We simply used sample mean, sample variance, and sample covariance as parameters in the t-distribution, and also simulated 2000 random sample. Result is shown at the bottom of Figure 4.8.

Despite the general advantage of the copula method that separate distributions can be used for the marginal distributions, the data at hand did not call for such a need. In fact, we ended up fitting t-distribution for marginals, and t-distribution for the copula. The comparison made in Figure 4.8 demonstrates the superiority of the copula method.

One unique characteristic seen in the observed dependency is that when movements are within (-0.05, 0.05) range, variability in the data seems to be very different from when movements are outside of the range. This apparent shift of dependency in the core and tail of the distribution is replicated in the simulation using t-Copula, but not at all in the bivariate t, as it is an elliptical distribution. It should also be pointed out, that simulation size of 2000 is chosen, which is about half of the observed sample size of 3852, so that simulated sample does not completely cover the observed dataset.
Figure 4.8: Observed vs Simulated using t-copula (above) and bivariate t-distribution. There are 3852 observations, and 2000 simulated points in each plot. Both models have similar standard deviations, correlation, and Kendall’s tau.
BIBLIOGRAPHY


