An Efficient Copula under Data Perturbations Across Stock Markets

Ivy Barley¹, Gabriel Asare Okyere¹, Henry Man’tieebe Kpamma², James Baah Achamfour¹, David Kweku¹, Godfred Zaachi¹

¹Department of Mathematics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
²Department of Statistics, Bolgatanga Polytechnic, Bolgatanga, Ghana

Abstract Economic trade amongst the various West African economies can either lead to mutual gains or losses. It is therefore important to assess the extent to which dependence amongst these countries can have on their economies. The linear correlation coefficient is normally used as a measure of dependence between random variables. However, there are some limitations when used for economic variables like the stock market; as they do not follow the elliptical distribution. Copulas, however are scale-free methods of constructing dependence structures amongst the stock markets, even in cases of data perturbations. The aim of this study is to assess the impact of data perturbations on the copula models. The maximum likelihood estimation method was the parameter estimation method used for the Archimedean copulas. The Clayton, Joe, Frank and Gumbel copulas were estimated. The Gumbel copula was the most robust copula in all the cases of data perturbations.

Keywords Copula, Dependence Structure, Data Perturbation, Joint Distribution

1 Introduction

The formation of regional bodies like Economic Community of West African States (ECOWAS) does not only have political implications to the countries in question, but economic ones as well. Trade amongst the various countries in the region can lead to mutual benefits or losses. It is imperative to assess the extent to which dependence amongst these countries influence their economies. According to Bekaert and Harvey [2], though the joint distribution of multivariate variables are usually assumed to follow the normal distribution, economic variables like the stock market index do not follow the normal distribution since they tend to be skewed, peaked and have extreme values. Though the linear correlation coefficient is normally used to assess the extent of dependence in joint normal distribution, Embrecht et.al.[9] highlighted the limitations and pitfalls of using it to study dependence amongst economic variables. Mahfoud and Michael [12] state that the copulas give a robust modelling tool that overcome the limitations of the Pearson correlation coefficient. Among the distinct classes of copulas, attention was set on the Frank, the Gumbel and the Clayton copula from the Archimedean class due to the different structures of dependence they consider. Also, from Chai [6], it has been established that linear Pearson correlation is not a fitting dependence measure for variables that do not follow the normal distributions.

1 Introduction

Chicheptiche [7] points out that bivariate stock returns are not elliptically distributed since a detailed test performed on the significance of Student models and generally elliptical models to explain the joint distribution of returns revealed that though Student copulas give a good estimation for stock pairs that have high correlation, systematic errors show up as the linear correlation coefficient between the pairs of stock reduce which rules out all the elliptical models. On the contrary, Aas [1] argues that linear correlation coefficient is the most utilized way of testing the dependence in economic variables and is just a measure of direct dependence implying an important measure of dependence if asset returns is represented well by elliptical distributions. Using linear correlation coefficient as a measure of dependence for distributions other than the elliptical ones may lead to conclusions that are misleading.

On the other hand Omelka et al. [13] state that one approach to model dependence structures is through copulas which is a way of capturing the dependence structures that exist in the joint distributions of variables. Measures of association like the Kendall’s tau or Spearman’s rho can be expressed as functions of the copula. where Bergsma et al. [3] add that the most prominent approaches to test whether two random variables that are ordinal are also independent is to employ Spearman’s rho and Kendall’s tau methods. Mahfoud and Michael [12] in stating the deficiency of Pearson correlation said it requires a more suitable dependence measure, which is the copula. Modeling economic variables using copulas has turned into an undeniably prominent method as it overcomes the constraints of the correlation especially in instances where losses are extreme and Caillault and Guegan [5] buttress the point by linking copulas to dependence of extreme. While linear correlation coefficient as a measure of dependence for distri-
butions other than elliptical one may lead to misleading conclusions. Aas [1] adds that other techniques must be used to capture dependency that exists amongst the random variables and one class of options are copula-based dependence measures where two parametric groups of copulas are considered; the copulas of normal mixture distributions and Archimedean copulas. Li et al.[11] therefore present a technique for building copula functions by joining the fields of convex sum and distortion, named Distorted Mix Method. This concept mixes distinctive copulas with margins that are distorted to build new copula functions. It empowers the individual to model the dependence structure of risks by factoring in central and extreme value parts independently. The tail dependence of a stated copula can then be changed to any predetermined level measured by function and coefficients of extreme values.

The copula has the ability of capturing tail dependence [10], an important concept in modeling bivariate financial data. One important attribute of the copula is that it accounts for the dependence structure that exist between the stock markets in question and just not the linear dependence [4]. The objective of this study is to assess the impact of data perturbation of the dependence structure that exist between the stock markets in question and just not the linear dependence [4]. The objective of this study is to assess the impact of data perturbation of the stock market index on the copula models.

2 Materials and Methods

There are various methods for the study of copulas. Among some of the methods considered in this paper are:

2.1 Frechet Bounds

The Frechet Bounds for any bivariate distribution \( W(x_1, x_2) \) is given by:

\[
\max M(x_1) + N(x_2) - 1, 0 \leq W(x_1, x_2) \leq \min M(x_1), N(x_2)
\]

where \( M(x_1) \) and \( N(x_2) \) are marginal distributions. The Frechet bounds play a vital role in the development of copulas.

2.2 Copula

The Cassell’s Latin Dictionary defines the copula as a bond, link or tie. There has been an ever growing interest in the applications of copulas in statistics. A copula is a function that links distributions that are multivariate in nature to their one-dimensional marginals. The marginals are uniform on the interval \((0,1)\). Copulas can be used to build multivariate stochastic models.

Definition 2.1 For every \( n \geq 2 \), an \( n \)-copula is a \( n \)-variate distribution function on \( I^n \) whose one-dimensional marginals are uniformly distributed on \( I \). Let \( M \) be a distribution function with \( n \) dimensions have univariate marginals \( M_1, M_2, \ldots, M_n \). Represent the range of \( M_k, A_k := M_k([0,1]) \) by \( A_k \). There exists a copula \( C \) such that for all \((y_1, y_2, \ldots, y_n) \in \mathbb{R}^n\)

\[
M(y_1, y_2, \ldots, y_n) = C(M_1(y_1), M_2(y_2), \ldots, M_n(y_n))
\]  

Such a \( C \) is uniquely defined on \( A_1 \times A_2 \times \ldots \times A_n \) and hence it is unique when \( M_1, M_2, \ldots, M_n \) are all continuous.

Theorem 2.1 (Sklar’s Theorem) [8] Let \( W \) represent joint distribution function which has univariate marginals \( M \) and \( N \). Then there exists a copula \( C \) such that for every \( x_1, y_2 \) in \( \mathbb{R} \),

\[
W(x_1, y_2) = C(M(x_1), N(y_2))
\]  

2.3 Tail dependence

Tail dependence is vital in modeling the dependence in cases of extreme events.

Definition 2.2 Let \( X = (X_1, X_2)^T \) be a 2-dimensional random vector with their univariate marginal distribution functions given as \( N_1 \) and \( N_2 \). The dependence coefficient of the upper tail of \( X \) is defined as:

\[
\lambda_U = \lim_{v \uparrow 1} \left( P(X_1 > N_1^{-1}(v)|X_2 > N_2^{-1}(v)) \right)
\]

The lower tail dependence coefficient is also defined as:

\[
\lambda_L = \lim_{v \downarrow 1} \left( P(X_1 < N_1^{-1}(v)|X_2 < N_2^{-1}(v)) \right)
\]

\( X \) is said to be lower tail dependent if and only if \( \lambda_L > 0 \) and is upper tail dependent if \( \lambda_U > 0 \). \( X \) is however upper tail independent if \( \lambda_U = 0 \) and lower tail independent if \( \lambda_L = 0 \).

2.4 Extreme Value Theory

Since estimating tail dependence is all about analyzing the extreme values of the distributions, we discuss the extreme value theory. The Generalized Extreme Value (GEV) distribution family is represented as:

\[
N(x) = \exp - \left( 1 + \alpha \left( \frac{x - \mu}{\beta} \right) \right)
\]

where \( (x)_+ = \max(x,0) \), \( \beta \in \mathbb{R}_+ \) and \( \alpha, \mu \in \mathbb{R} \). Also, \( \mu \) is the location parameter, \( \beta \) is the scale parameter and \( \alpha \) is the shape parameter.

2.5 Measures of Association

Kendall’s \( \tau \) and Spearman’s \( \rho \) measure concordance, which is a form of dependence.

2.5.1 Spearman’s \( \rho \) with respect to Copulas

The Spearman’s \( \rho \) can be computed with respect to copulas and this is represented by:

\[
\rho_c = 12 \int_0^1 \int_0^1 (C(u_a, v_b) - u_a v_b) \, du \, dv
\]

2.5.2 Kendall’s \( \tau \) with respect to Copulas

The Kendall’s \( \tau \) can be computed with respect to copulas and this is represented by:

\[
\tau_c = 4 \int_0^1 \int_0^1 C(u_a, v_b) \, dC(u_a, v_b) - 1
\]

The relationship between both measures of concordance can be observed by this theorem:
Theorem 2.2 Let $X$ and $Y$ be continuous random variables and let $\rho$ and $\tau$ represent Spearman's $\rho$ and Kendall's $\tau$. Then

$$-1 \leq 3\tau - 2\rho \leq 1$$ (9)

2.6 Families of Copulas

Though there are a variety of families of copulas, the most important ones are the Archimedean and elliptical families.

2.6.1 Elliptical Copula

The random vector $X = (X_1, \ldots, X_d)$ has a multivariate elliptical distribution, represented as $X \sim El(\mu, \Sigma, \psi)$, if for $x = (x_1, \ldots, x_d)$ and the form of its characteristic function is

$$\varphi(x, \mu, \Sigma) = \exp(iX'\mu)\psi_d\left(\frac{1}{2}x'\Sigma x\right)$$

where $\mu$ is a vector, $\Sigma$ is a symmetric positive-definite matrix and $\psi_d(t)$ is the characteristic generator function.

There are two types of elliptical copulas namely Gaussian copula and Student t-copula. The Gaussian copula is given by

$$C_\rho(u_1, u_2) = \Phi_\Sigma(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$$ (10)

where $\Sigma$ is a $2 \times 2$ covariance matrix, $\Phi$ is the cumulative distribution function of a standard normal distribution and $\Phi_\Sigma$ is the cumulative distribution function of the bivariate normal distribution which has mean 0 and covariance matrix $\Sigma$.

The Student t-copula is also given by

$$C_v, \Sigma(u_1, u_2) = t_v, \Sigma(t_v^{-1}(u_1), t_v^{-1}(u_2))$$ (11)

where $\Sigma$ is a correlation matrix, $t_v$ is the cumulative distribution function of the one dimensional $t_v$-distribution and $\Sigma$ is the cumulative distribution function of the multivariate $t_v, \Sigma$-distribution.

2.6.2 Archimedean Copulas

An $n$-dimensional copula is called Archimedean if it is of the form:

$$C(u_1, u_2, \ldots, u_n) = \psi^{-1}\left(\psi(u_1) + \psi(u_2) + \cdots + \psi(u_n)\right)$$ (12)

where $u_1, u_2, \ldots, u_n \in [0, 1]$ and $\psi$ is a decreasing and continuous function mapping from $[0, 1]$ into $[0, \infty]$ and this is called the generator of the copula. It is a strict generator however when $\psi(0) = \infty$.

2.7 Maximum Likelihood Estimation

Maximum Likelihood Estimation is one of the methods used to obtain the copula parameter estimates by maximizing the log likelihood function. It is a parametric technique that is useful in fitting copulas to data. The maximum likelihood is asymptotically unbiased and efficient. The maximum likelihood function, $L$ can be decomposed into:

$$L = L_c + \sum_{i=1}^{d} L_i$$ (13)

### Table 1. The Jarque-Bera Test

<table>
<thead>
<tr>
<th>Stock Market</th>
<th>OD</th>
<th>PT1</th>
<th>PT2</th>
<th>PT3</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE</td>
<td>111.73</td>
<td>104.95</td>
<td>137.73</td>
<td>383.53</td>
</tr>
<tr>
<td>NGSE</td>
<td>470.02</td>
<td>368.27</td>
<td>515.98</td>
<td>572.38</td>
</tr>
<tr>
<td>BRVM</td>
<td>73.76</td>
<td>65.89</td>
<td>58.29</td>
<td>58.24</td>
</tr>
</tbody>
</table>

where

$$L_c = \log c\left(F_1(x_1^{(j)}), F_2(x_2^{(j)}), \ldots, F_d(x_d^{(j)})\right)$$ (14)

is the log-likelihood from the dependence structure represented by the copula $C$ and

$$L_i = \log f_i(x_i^{(j)})$$ (15)

are the log-likelihoods from the dependence of each marginal.

The data is made up of the stock market index of Economic Community Of West African States (ECOWAS) countries and was derived from African Financial Markets. The trading period was from December, 2014 to January 2017 and consists of 498 complete trading days. The countries include Ghana (GSE), Nigeria (NGSE) and BRVM which is made up of the Francophone countries (Benin, Cote D’Ivoire, Burkina Faso, Togo) where the trading were done simultaneously. The sampled data was sufficient for a robust inference to be made for the study.

2.8 Summary Statistics

From the GSE, NGSE and BRVM of the Original Data, there are no outliers in the data set. The concept of outliers is important in this research as we want to ascertain the impact that the presence or absence of outliers will have on the choice of copula. A perturbation of outliers were however introduced at 5%, 10% and 20% at both the minimum and maximum values of each stock market index, subsequently represented as OD, PT 1, PT 2 and PT 3 respectively.

2.9 The Jarque-Bera Test

The Jarque-Bera test statistic were computed for the stock markets and decision was made at 5% significance level. Since the Jarque-Bera test statistic asymptotically follows the chi-squared distribution, the computed statistic will be compared against $\chi^2$ (with 2 degrees of freedom) = 5.99. Due to the fact that the Jarque-Bera statistic for all the stock markets are significantly greater than 5.99, it can be concluded that the data set does not follow the normal distribution. Hence, modeling them with the elliptical distribution would not be appropriate.

2.10 Parameter Estimation

For this research, the maximum likelihood estimation method was used to estimate the parameters of the Clayton, Joe, Frank and Gumbel copulas.
2.10.1 Maximum Likelihood Estimates

The maximum likelihood estimation procedure gives us the parameter estimates for dependence of each copula, as shown in Tables 2, 3 and 4. In all the various stock pairs, the following can be observed:

1. The Clayton copula parameter increases by a scalar multiple of about 2.
2. The Joe copula parameter decreases by about half.
3. The Frank copula increases or decreases haphazardly.
4. The Gumbel copula parameter increases insignificantly.

From the estimates, the Gumbel copula is the most robust copula with or without the presence of outliers.

2.11 Tail Dependence Analysis

From the tail dependence tables of all stock pairs, the lower and upper tail dependence coefficients show a similar trend. Of all the tail dependence coefficients estimated, the Gumbel copula models best show instances of economic boom as the stock market pairs tend to perform best when both are in their upper extreme values. This is followed by the Joe copula.

2.12 QPlots Of Copulas

The QPlots (Fig. 1, 2, 3 and 4) are used to graphically show the tail dependence coefficients of each stock market pair.
due to the tail dependence figure of 0.40. The upper section captures the dependence structure of the two stock markets.

### 2.13 Probability Density Functions of Copulas

The probability density function of each copula of the original data are represented. Probability density functions are important in the study of copulas to determine the relationship that exist between the stock pairs.

### 2.14 Model Selection

When the parametric estimation method is used, the Bayesian Information Criterion (BIC) is more appropriate for the model selection. Taking into account the pair-wise relationships between the stock markets, it can be seen from Table 8 that the Gumbel copula has the least BIC for GSE and NGSE per the data used. The implication is that the Gumbel copula is the best copula to model the dependence structure between Ghana and Nigeria. For GSE and BRVM, it can also be seen that the Gumbel copula models best the dependence structure between GSE and BRVM. Finally for BRVM and NGSE, the Joe copula models best the dependence structure that exist between them.

### 2.15 Estimated Copula Models

The following are the copula models for Ghana and Nigeria based on the Bayesian Information Criterion (BIC): With the association parameters obtained from the copula models from Table 11, the Kendell’s tau correlation measure for the Frank, Clayton and Gumbel copula are estimated to be 0.47, 0.853, 0.97 respectively; suggesting a positive correlation between the Ghanaian and Nigerian Stock markets. Given any values of $u$ and $v$, the copula can be estimated for each stock market pair. This helps to ascertain the relationship that exist between the stock market pairs; and more generally their dependence structures.

### 3 Conclusions

All the stock market pairs under study in this work did not follow the normal distribution. This was ascertained from the Jarque-Bera test statistic. Hence, elliptical copulas were not appropriate to model the dependence structures that exist between the stock pairs. All the stock market pairs were positively skewed, implying that the longer tails were on the positive sides of the peak. Positive kurtosis statistics also showed that the stock markets had fatter and longer tails. To assess the im-
The presence of outliers on the copulas for each stock market pairs, outliers were introduced at 5 percent, 10 percent and 20 percent, which were subsequently represented as PT 1, PT 2 and PT 3 respectively. The original data was represented by OD. The maximum likelihood estimation (MLE) procedure was used to estimate the dependence parameters for each copula. For all the copula models under study, it was seen that the Gumbel copula was the most robust; as the presence of outliers at each level did not cause its copula parameter to change significantly. The tail dependence coefficients helps to compare the extent of tail-heaviness in the stock market pairs. Both lower tail dependence coefficients and upper tail dependence coefficients were estimated. The Clayton copula was more correlated at the lower extreme values implying that the stock market pairs were less likely to gain together and more likely to lose together. This depicted economic losses. The Gumbel copula however showed more upper tail dependence which means that the stock market pairs are more likely to gain together than lose together. Hence the Gumbel copula is the most robust.

### 3.1 Recommendations

This research work sought to assess the effect of data perturbations on the choice of copulas. The parametric method of estimation was used at all levels. It is however recommended that non-parametric and semi-parametric estimation methods be used in subsequent work. This can be used to compare if the effect of data perturbation is influenced by the choice of estimation methods. Future work can also consider the time-varying dependence structure at each level of perturbation of the stock market by using the conditional copula. This may include evaluating and comparing time series density models at each level of data perturbation.

### REFERENCES


