

Copulas and Correlation in Statistical Risk Theory

Capstone Project by Sara Rat

Supervisor: Dr.Dongming Wei, Second Reader: Dr.Zhenisbek Assylbekov

1 Introduction

The Financial Risk Management (FRM) aims to identify, measure and manage risks in different sectors. One of the core things during such operations is measuring different dependencies. Linear correlation is known as one of the most popular measure of dependence, however it is known that it is a reasonable measure of dependence only when variables are Normally distributed, but this is not the case with credit and portfolio risks, therefore other measures of dependency are needed. This paper presents a *Copula* function for bivariate case as one of the tools to analyze dependencies in portfolio risk management.

Copulas were first introduced by Sklar in 1959 [8], and in 1999 they were studied in financial context for the first time by Embrechts et al. in 1999 [4].

Motivated by the copula analysis of European stock portfolios [6], this paper aims to analyze portfolio consisting of Asian S&P Asia 50 and S&P BSE 100 indices, and apply copula to this portfolio.

2 Theory

2.1 Dependence measures

In this section, different dependence measures will be discussed, namely Pearson linear correlation, rank correlation and tail dependence. Further it will be discussed that the classic approach used to measure dependence, which is the Pearson linear correlation has some drawbacks which may cause problems when data is analyzed. Since the research focuses on bivariate copulas, only bivariate cases of dependence measures will be discussed.

2.1.1 Pearson linear correlation

Pearson linear correlation coefficient r measures direction and strength of a linear relationship between two random variables.

$$r = \frac{Cov[X, Y]}{\sqrt{Var(X)Var(Y)}} \quad (1)$$

Where $Cov[X, Y]$ is a covariance between random variables X and Y , and $Var(X)$ and $Var(Y)$ are variances of two random variables, and $-1 \leq r(X, Y) \leq 1$

However this correlation coefficient r has a number of pitfalls when the data is analyzed [4, 5]:

- It is a reasonable measure of dependence only when variables are Normally distributed.
- It is a scalar measure of dependence; doesn't tell much about the dependence structure of risks.
- A correlation of zero does not indicate independence of risks.
- Perfect positively dependent risks do not necessarily have a correlation of 1.
- perfect negatively dependent risks do not necessarily have a correlation of -1.
- Correlation is not invariant under strictly increasing transformations of the risks.
- Correlation is only defined when the variances of the risks are finite.

2.1.2 Kendall's tau

Since the previous section showed that the Pearson linear correlation has a number of disadvantages, it is needed to consider other ways of measuring dependence of random variables. One of such measures is the Kendall's rank correlation. Before introducing this important measure, it is needed to define the concept of *concordance* [8].

Definition 2.1. Let (x_i, y_i) and (x_j, y_j) be two observations from a vector (X, Y) of continuous random variables. Then, (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, (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 $y_i < y_j$ [8].

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.

Definition 2.2. Let $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ denote a random sample of n observations from a vector (X, Y) of continuous random variables. There are $\binom{n}{2}$ 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

$$\tau = \frac{(c - d)}{n(n - 1)/2} \quad (2)$$

Kendall's tau can be generally defined as the difference between the probability of concordance and probability of discordance for a pair of observations (x_i, y_i) and (x_j, y_j) that is chosen randomly from the sample [8].

Copulas play important role in concordance and measures of association such as Kendall's τ [8]. Let X and Y be continuous random variables, then Kendall's τ can be expressed as [2]:

$$\tau(X, Y) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 \quad (3)$$

Where $C(u, v)$ is the copula of the bivariate distribution function of X and Y . Kendall's tau can be considered as a measure of monotonic dependence, whereas linear correlation is a measure of linear dependence only. In [7] cited in [6] it is concluded that it is considered better to use Kendall's tau than Pearson linear correlation to describe the dependence, because it is invariant under monotonic non-linear transformations of the underlying variables. However, Kendall's tau is also a scalar measure of dependence, and it is needed to have some other measurements to describe the dependence structure of the data in more details. One of such measurements is a *tail dependence* discussed in the following subsection.

2.1.3 Tail dependence

Tail dependence describes the amount of dependence in the tail of bivariate distribution. This dependence measure looks at concordance between extreme values of the continuous random variables X and Y . Geometrically, it focuses on the upper and lower quadrant tails of the joint function [2]. Nelsen [8] defines the parameters of upper and lower tail dependence in the following way:

Definition 2.3. Let X and Y be continuous random variables with distribution functions F and G , respectively. The *upper tail dependence parameter* λ_U is the limit (if it exists) of the conditional probability that Y is greater than 100t-th percentile of G , given that X is greater than the 100t-th percentile of F as t approaches 1, i.e.

$$\lambda_U = \lim_{t \rightarrow 1^-} P[Y > G^{-1}(t) | X > F^{-1}(t)] \quad (4)$$

Similarly, the *lower tail dependence parameter* λ_L is the limit (if it exists) of the conditional probability that Y is less than or equal to the 100t-th percentile of G given that X is less than or equal to the 100t-th percentile of F as t approaches 0, i.e.

$$\lambda_L = \lim_{t \rightarrow 0^+} P[Y \leq G^{-1}(t) | X \leq F^{-1}(t)] \quad (5)$$

These parameters are nonparametric and depend only on the copula of X and Y , this can be seen from the following theorem given by Nelsen [8].

Theorem 1. Let $X, Y, F, G, \lambda_U,$ and λ_L be as in Definition 2.1, and let C be the copula of X and Y . If the limits (2) and (3) exist, then

$$\lambda_U = 2 - \lim_{t \rightarrow 1^-} \frac{1 - C(t, t)}{1 - t} \quad (6)$$

$$\lambda_L = \lim_{t \rightarrow 0^+} \frac{C(t, t)}{t} \quad (7)$$

The variables X and Y are said to be asymptotically independent if $\lambda_U(X, Y) = \lambda_L(X, Y) = 0$.

2.2 Copula Definition and Properties

Definition 2.4. A *copula function* links n univariate marginal distributions to a full multivariate distribution resulting in a joint distribution function of n standard uniform random variables [5]. In other words, copulas use individual cumulative distribution functions (CDFs) as inputs and return joint probabilities. In our study we focus only on the bivariate case.

Assume X and Y are random variables. Then, their joint CDF H is given by

$$H(x, y) = C(F(x), G(y)) \quad (8)$$

Where $C(u, v)$ is a copula, F and G are marginal distribution functions. Copula is a multivariate distribution function from the unit d -cube $[0, 1]^d$ to the unit interval $[0, 1]$ which satisfies the following properties discussed in [8, 6]:

- $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i \forall i \leq d$ and $u_i \in [0, 1]$
- $C(u_1, \dots, u_d) = 0$ if $u_i = 0 \forall i \leq d$
- C is *d-increasing*. This property ensures that the joint probability will not be negative, since the volume (C) of any d -dimensional interval is non-negative.
- For every copula $C(u_1, \dots, u_d)$ there are the Fréchet Bounds:

$$\max\left[\sum_{i=1}^d u_i + 1 - d, 0\right] \leq C(u) \leq \min[u_1, \dots, u_d] \quad (9)$$

2.3 Sklar's Theorem

One of the fundamental theorems of this topic is the *Sklar's Theorem*. Chan and Wong [1] state it in the following way:

Theorem 2. Let F be a joint distribution function with marginal distribution functions F_1, \dots, F_n . Then there exists a copula C such that for all $x_1, \dots, x_d \in R = [-\infty, \infty]$,

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (10)$$

- If F_1, \dots, F_d are continuous, then C is unique; otherwise C is uniquely determined on $\text{Ran}F_1 \times \dots \times \text{Ran}F_n$, where $\text{Ran}F_j$ denotes the range of F_j .
- if C is a copula and F_1, \dots, F_d are univariate distribution functions, then F is defined by (3) is a joint distribution function with marginal distribution functions F_1, \dots, F_d .

2.4 Copula Families, Archimedean Copulas

There are two main Copula families: Archimedean copulas (Clayton, Gumbel, Frank) and elliptical copulas (Gaussian, Student t). The most popular copulas used in finance/risk management are: Gaussian, Student-t, Clayton and Gumbel [3]. This paper focuses only on bivariate Archimedean copulas, and the most important copulas of this class will be discussed in details, namely Clayton, Gumbel, and Frank copulas. The following definitions in this section are given by Cheburini, Luciano and Vecchiato [2].

Definition 2.5. *Archimedean copulas* may be constructed using a function $\phi : I \rightarrow R^+$, continuous, decreasing, convex and such that $\phi(1) = 0$. Such a function ϕ is called a **generator**. It is called a strict generator whenever $\phi(0) = +\infty$. The **pseudo-inverse** of ϕ must also be defined, as follows:

$$\phi^{[-1]}(v) = \begin{cases} \phi^{-1}(v), & 0 \leq v \leq \phi(0) \\ 0, & \phi(0) \leq v \leq +\infty \end{cases} \quad (11)$$

The pseudo-inverse is such that, by composition with the generator, it gives the identity, as ordinary inverses do for the functions with domain and range :

$$\phi^{[-1]}(\phi(v)) = v$$

for every $v \in I$. In addition, it coincides with the usual inverse if ϕ is a strict generator.

Definition 2.6. Given a generator and its pseudo-inverse, an *Archimedean copula* C^A is generated as follows:

$$C^A(v, z) = \phi^{[-1]}(\phi(v) + \phi(z)) \quad (12)$$

If the generator is strict, the copula is said to be a strict Archimedean copula.

2.4.1 One-parameter Archimedean copulas

This paper studies one-parameter Archimedean copulas, which are constructed using a generator $\phi_\theta(t)$, indexed by the (real) parameter θ . It is possible to get a subclass or family of Archimedean copulas by choosing a particular generator. This paper focuses on three well-known families and their generators, which are discussed in [2] and shown in Table 1.

Copula	$\phi_\theta(t)$	range of θ	C(u,v)
Gumbel	$(-\ln t)^\theta$	$[1, +\infty]$	$\exp(-[(-\ln u)^\theta + (-\ln v)^\theta]^{\frac{1}{\theta}})$
Clayton	$\frac{1}{\theta}(t^{-\theta} - 1)$	$[-1, 0) \cup (0, +\infty)$	$\max[(u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}, 0]$
Frank	$-\ln \frac{\exp(-\theta t) - 1}{\exp(-\theta) - 1}$	$(-\infty, 0) \cup (0, +\infty)$	$\frac{1}{\theta} \ln(1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1})$

Table 1: Some One-parameter Archimedean Copulas

As it was discussed in subsections 2.1.2 and 2.1.3, copulas are closely associated with Kendall's tau and tail dependence. The following table gives formulas for calculation of these dependence measures for some one-parameter Archimedean copulas [8].

Copula	Kendall's tau	Upper tail	Lower tail
Gumbel	$1 - \theta^{-1}$	$2 - 2^{\frac{1}{\theta}}$	0
Clayton	$\frac{\theta}{\theta+2}$	0	$2^{-\frac{1}{\theta}}$
Frank	$1 + 4[D_1(\theta) - 1]/\theta$	0	0

Table 2: Kendall's tau and Tail dependence of some Archimedean copulas

Remark. In the table above, *Debye* function was used in order to find Kendall's tau for Frank copula, and it is defined as:

$$D_n(x) = \frac{n}{x^n} \int_0^x \frac{t^n}{e^t - 1} dt$$

2.4.2 Copula Parameter Estimation

It was discussed in the previous subsection that Archimedean copulas are constructed using a generator function $\phi_\theta(t)$, which is indexed by a parameter θ . This parameter is a *copula parameter*, and it should be estimated from the data. One of the most widely used methods of estimation used to estimate this parameter is *exact maximum likelihood method*. This method is described in details by Cheburini, Luciano, and Vecchiato [2]. Before introducing this method, it is important to review some canonical representation for two random variables X_1 and X_2 :

$$f(x_1, x_2) = c(F_1(x_1), F_2(x_2))f_1(x_1)f_2(x_2)$$

where

$$c(F_1(x_1), F_1(x_2)) = \frac{\partial^2(C(F_1(x_1), F_2(x_2)))}{\partial F_1(x_1)\partial F_2(x_2)}$$

is the second mixed partial derivative of the copula C, c is the copula density and f is the standard univariate probability density function. It is stated in [2] that a statistical modeling problem for copulas consists of two steps:

- marginal distributions' identification
- the appropriate copula function definition

One of the major steps in this procedure is a copula parameter estimation. Let $D = (x_{1t}, x_{2t})_{t=1}^T$ be the sample data matrix. Then, the expression for the log-likelihood function is:

$$l(\theta) = \sum_{t=1}^T \ln c(F_1(x_{1t}), F_2(x_{2t})) + \sum_{t=1}^T \sum_{j=1}^2 \ln f_j(x_{jt}) \quad (13)$$

where θ is the set of all parameters of both marginals and the copula. Thus, by knowing marginal p.d.f.s and given a copula, the function (11) can be written, and then by maximization, the maximum likelihood estimator can be found:

$$\hat{\theta}_{MLE} = \arg \max_{\theta} l(\theta) \quad (14)$$

This estimation is usually done by numerical maximization method. In this paper we assume that the maximum likelihood estimator exists, is consistent and asymptotically normal as in [2]. Moreover,

$$\sqrt{T}(\hat{\theta} - \theta_0) \sim N(0, J^{-1}(\theta_0))$$

where J is the Fisher's information matrix and θ_0 is the usual true value.

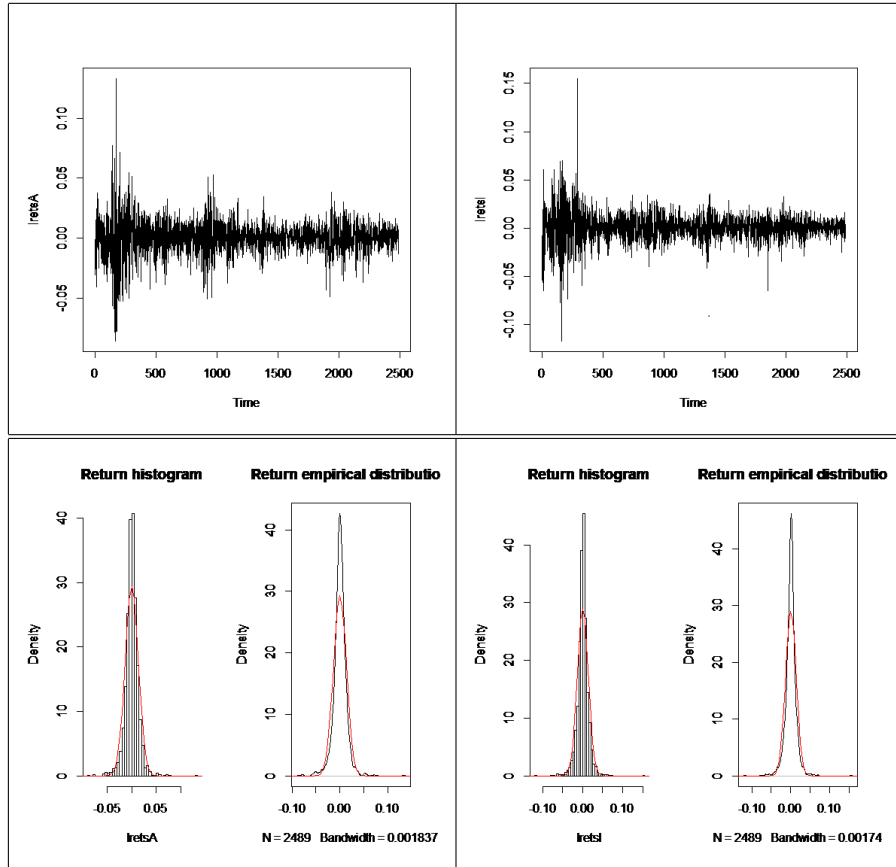
3 Application

3.1 Data description

The portfolio that is analyzed consists of two indices: SP BSE 100 (India) and SP Asia 50 (Hong Kong, Korea, Singapore and Taiwan), since it is assumed that these countries represent general economy of Asia. The daily closing prices of these indices were obtained from SP Dow Jones Indices database [11] of the period from 29.02.2008 to 23.03.2018 were used. The linear correlation test for these indices showed that correlation coefficient is 0.7903507, which means that indices seem to be highly correlated.

The figure above shows time plots and distribution plots of the SP Asia 50 and SP BSE 100 log-return series. It can be seen that on both time plots large (small) returns are followed by small (large) returns. Also, distribution plots show that both indices deviate from normal distribution (normal distribution for given mean and standard deviation is shown with red lines). In addition, it can be observed that both distribution plots have high peak around the mean. Table 3 shows more detailed information about the log-returns of both indices.

Table 3 shows that both log-returns have positive meaning and large kurtosis (it can also be visually seen from distribution plots of the indices), and SP Asia



50 is slightly left-tailed, it can be seen from negative skewness, whereas SP BSE 100 is right-tailed. In order to check the normality, the Jarque-Bera (JB) test was used, and the results can be seen from Table 2. It can be seen that p-value of JB-test is very small, which means that we strongly reject the hypothesis that states that data is normally distributed for both indices. In addition, Ljung-Box test for squared log returns for first 8 lags was conducted and it showed that there is a serial correlation at 1 % significance level in both indices. Thus, the fact that log-returns are not normally distributed and have some serial correlation, ARMA-GARCH model should be used to filter our data and give serially independent data.

3.2 Modeling the marginal distributions

As it was stated in the previous section, log-return of indices should be modeled using ARMA-GARCH model, where error terms will have student-t distribution. Before doing this modeling, the Dickey-Fuller test was conducted to see whether

Stock Index	Min	Median	Mean	Max	Std	Skewness	Kurtosis
Asia	-0.086	0.0004	0.0001	0.133	0.013	-0.021	8.318
India	-0.117	0.0007	0.0003	0.155	0.014	0.077	11.637

Table 3: Statistics of SP Asia 50 and SP BSE 100

Stock Index	JB	JB p-value	LB for squared returns	LB p-value
Asia	7191.4	$< 2.2e - 16$	1387.1	$< 2.2e - 16$
India	14074	$< 2.2e - 16$	536.99	$< 2.2e - 16$

Table 4: Jarque-Bera and Ljung Box tests to test normality of log returns and for the normality of ARCH effect in squared log returns respectively

our log-returns are stationary. The test showed that both Asian and Indian log returns are stationary, p-value $\hat{=}$ 0.01. Then, the ARMA(p,q)-GARCH(m,s) model is given by:

$$X_{i,t} = \mu_i + \sum_{j=1}^p \phi_j X_{i,t-j} + \sum_{k=1}^q \theta_{i,k} \epsilon_{i,t-k} + \epsilon_{i,t} g_{i,t} \quad (15)$$

$$\epsilon_{i,t} = \sigma_{i,t} \eta_{i,t} \quad (16)$$

$$\sigma_{i,t}^2 = \alpha_{i,0} + \sum_{j=1}^m \alpha_{i,j} \epsilon_{i,t-j}^2 + \sum_{k=1}^s \beta_{i,k} \sigma_{i,t-k}^2 \quad (17)$$

Where $i=(\text{SP 50 Asia, SP BSE 100})$, and

$$\alpha_{i,0} > 0, \alpha_{i,j} \geq 0, \sum_{j=1}^{\max(m+s)} (\alpha_{i,j} + \beta_{i,j}) < 1 \quad (18)$$

The ARMA(p,q)-GARCH(m,s) model's $\mu, \phi, \theta, \alpha, \beta$ parameters are estimated using conditional likelihood approach. The function is the following:

$$l(\theta_i) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{t=2}^n \log \sigma_{i,t}^2 - \frac{1}{2} \sum_{t=1}^n \frac{\epsilon_{i,t}^2}{\sigma_{i,t}^2} \quad (19)$$

Where $\theta_i = \mu_i, \phi_{i,1}, \dots, \phi_{i,p}, \theta_{i,1}, \dots, \theta_{i,q}, \alpha_{i,0}, \dots, \alpha_{i,m}, \beta_{i,1}, \beta_{i,s}$ Now, the maximum likelihood estimator (MLE) maximizes the log likelihood and is given by

$$\theta'_i = \text{argmax}_{\theta_i} l(\theta_i) \quad (20)$$

Usually, MLE is found by numerical optimization approach.

In order to identify the lags of ARMA and GARCH (p,q, r,s values), PACF and ACF plots were empirically analyzed. Log-returns were used as inputs for ARMA lags' search, and squared log-returns were used for GARCH, but the

plots used were the same for both of them (PACF and ACF). The plots of PACF and ACF can be found in the Appendix. After empirical analysis and comparison of p-values of LB-test, the most appropriate lags were chosen. So, the optimal model for Asia is ARMA(3,1)-GARCH(1,1) with p-value= 0.01556, and ARMA(1,1)-GARCH(1,1) model was chosen for India, with p-value=0.03795. With these p-values LB-test fails to reject, hence there is no serial correlation in data now at 1 % significance level. These independent values can be further used to structure the marginal distributions of SP Asia 50 and SPE BSE 100.

3.3 Copula selection

The *copula* and *VineCopula* libraries were used in R to work with our data. After obtaining serially independent data, standardized residuals from ARMA-GARCH model were transformed to uniforms, using student-t distributions. These uniform series were used as inputs for bivariate Archimedean Copulas mentioned in Section 2.4.1. Copula parameter was estimated for each of these copulas, and maximum likelihood estimation method was used for that. The results can be seen from the following table:

Copula	Estimated Parameter	AIC	BIC
Clayton	1.24	-290.59	-284.77
Gumbel	1.81	-421.42	-415.6
Frank	4.09	-175.64	-169.82

Table 5: Estimation of copula parameter for SP Asia 50 and SP BSE 100

It was stated in Section 1. that this study was motivated by the copula analysis of European stock portfolios [6], and the following table shows the results obtained by the study in [6]:

Copula	Estimated Parameter	AIC	BIC
Clayton	3.718	-4591.805	-4585.725
Gumbel	3.840	-5445.510	-5439.430
Frank	13.770	-5301.488	-5295.407

Table 6: Estimation of copula parameter for DAX-30 and CAC-40 indices (Result of the Master's thesis)

Copula selection is done by using *BIC* and *AIC* criteria [9]. *BIC* is a Bayesian information criterion, whereas *AIC* is an Akaike information criterion, defined as:

$$AIC = 2k - 2\ln(L') \quad (21)$$

$$BIC = \ln(n)k - 2\ln(L') \quad (22)$$

Where L' is a maximized value of likelihood function, n is a number of data points, k is a number of parameters estimated by the model. Generally, BIC works better with large samples, whilst AIC is considered to be good in working with small samples [10]. The model with the lowest AIC and BIC is considered to be the most appropriate. Since Table 5 shows that that Gumbel copula has the lowest both AIC and BIC, it can be concluded that Gumbel copula is the most appropriate for our data among one-parameter Archimedean copulas. Table 6 also shows that Gumbel copula has the lowest both AIC and BIC, so the results of the master's thesis [6] shows the same result for different data. One of the main properties of Gumbel copula is a strong upper tail dependence. The plot of the copula can be seen from the following figure:

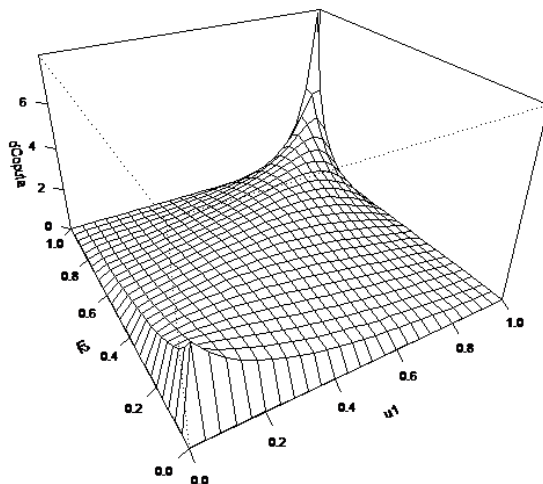


Figure 1: Gumbel copula density plot

3.3.1 Dependence measures

It was stated in subsections 2.1.2 and 2.1.3, copulas are highly associated with Kendall's tau and with tail dependence, and Table 2 describes the relationship between them in details. Since, our estimated copula is a Gumbel copula ($\theta = 1.82$), the estimation of Kendall's tau and tail dependence is straightforward. The estimated values are shown in Table 7. Since Kendall's tau is not close to 1, we cannot conclude that there is a very strong correlation between two stock indices, however there is some positive correlation. As for tail dependence, Gumbel copula cannot capture lower tail dependence, therefore it is equal to 0. In contrast, it perfectly captures upper tail dependence, and in our case it is

Kendall's tau	Upper tail	Lower tail
0.45	0.53	0

Table 7: Estimated Kendall's tau and tail dependence based on the estimated copula

0.53, which means that there is some upper tail dependence, since it is greater than 0, but it is not very strong.

4 Conclusion

This paper discussed pitfalls of using linear correlation in dependence measures, and concluded that using linear correlation in dependence measures can be insufficient. Then, copula function and one of its families named Archimedean copulas was introduced, and it was discussed that copula can be used to improve measurement of dependency. Afterwards, the portfolio consisting of SP Asia 50 and SP BSE 100 stock indices was analyzed. To deal with non-normality and heavy tails and serial correlation, ARMA-GARCH model was used, which filtered data and provided serially independent innovations in the end. Then, *copula* and *VineCopula* libraries of R were used to fit one-parameter Archimedean copulas to the data, and the AIC and BIC criteria showed that Gumbel copula is the most appropriate to use for this data. Then, Kendall's tau and tail dependence for the SP Asia 50 and SP BSE 100 stock indices were estimated by successful construction of the copula function. It was concluded that there is some positive correlation between two stock indices, and they have some upper tail dependence. The lower tail dependence was not found, because Gumbel copula does not capture the lower tail dependence. The initial aim of this paper was to use copulas to give more detailed information about the dependence structure of the data, and now it is known that different copulas can be used to fit the data, and then they can be used to estimate important dependence measures (Kendall's tau and tail dependence in our case).

5 Suggestions for further research

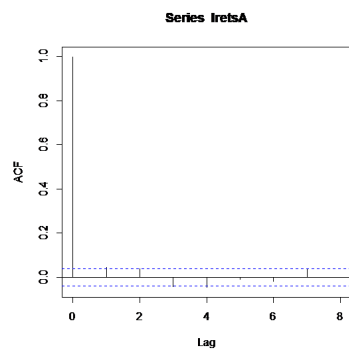
Since this project focuses only on bivariate one-parameter copulas, there is a big field for doing research for multivariate multiple-parameter copulas, because in real life portfolios consist of more than two indices. For example, this study can be broadened by applying other copulas to this data, for example, Elliptical copulas.

References

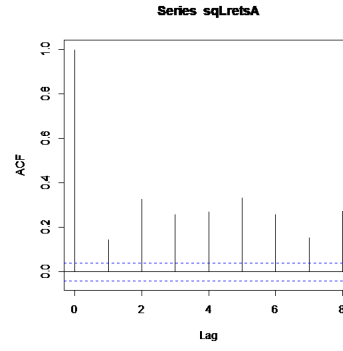
- [1] Chan, N. H., and Wong, H. Y. (2013). Handbook of Financial Risk Management: Simulations and Case Studies (Vol. 12), John Wiley and Sons.
- [2] Cherubini, U., Luciano, E., and Vecchiato, W. (2004). Copula methods in finance, John Wiley and Sons.
- [3] Crook, Jonathan, and Moreira (2011), Checking for asymmetric default dependence in a credit card portfolio: A copula approach *Journal of Empirical Finance* 18.4: 728-742.
- [4] Embrechts, P., McNeil, E., Straumann, D. (1999), Correlation: pitfalls and alternatives, Risk Magazine.
- [5] Habiboallah, Farzia (2007), Modeling dependencies in financial risk management
- [6] Mahfoud, M., and M. Michael (2012), Bivariate Archimedean copulas: an application to two stock market indices, BMI Paper
- [7] McNeil, A., Fre, R. and Embrechts, P. (2005), *Quantitative Risk Management: Concepts, Techniques, and Tools*, Princeton University Press
- [8] Nelsen, R. B. (2007), An Introduction to copulas, *Springer*, New York
- [9] Parra Palaro, H. and Koodi Hotta, L., 2006, Using Conditional Copula to Estimate Value at Risk, *Journal of Data Science*, 4, 93-115
- [10] Shumway, Robert H. and Stoffer, David S., 2011, Time Series Analysis and Its Applications, Springer, New York
- [11] "S&P Dow Jones Indices." SP Dow Jones Indices. Accessed March 15, 2018. <https://us.spindices.com/>.

6 Appendix

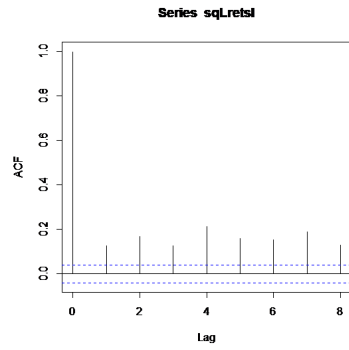
There are ACF and PACF plots of log-returns of SP 50 Asia and SP BSE 100



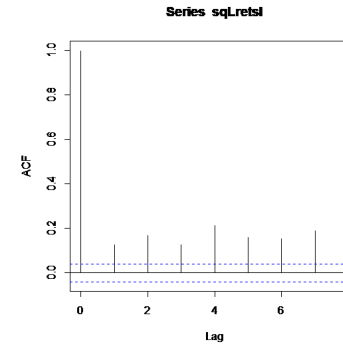
(a) ACF of log returns of SP 50 Asia



(b) ACF of squared log returns of SP 50 Asia

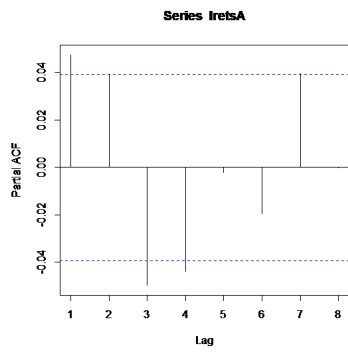


(c) ACF of log returns of SP BSE 100

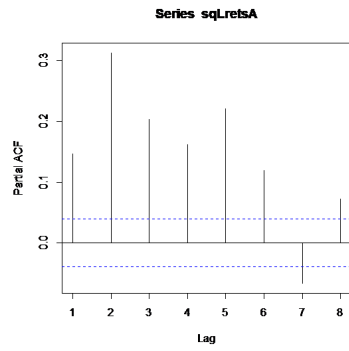


(d) ACF of squared log returns of SP BSE 100

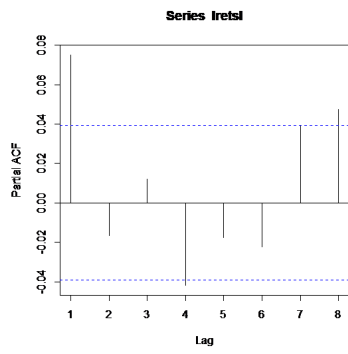
Figure 2: ACF plots



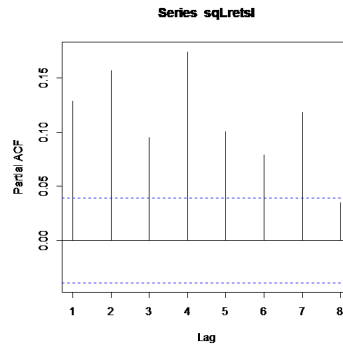
(a) PACF of log returns of SP 50 Asia



(b) PACF of squared log returns of SP 50 Asia



(c) PACF of log returns of SP BSE 100



(d) PACF of squared log returns of SP BSE 100

Figure 3: PACF plots

R Script

```
data<- read.table("maindata.txt", header=TRUE, sep="\t",\ \ stringsAsFactors=FALSE)\ \
data\ \
data$Asia\ \
attach(data)\ \
mean(Asia)\ \
mean(India)\ \
n<- length(Asia)\ \
lretsA <- log(Asia[-1]/Asia[-n])\ \
head(lretsA)\ \
summary(lretsA)\ \
lretsI <- log(India[-1]/India[-n])\ \
summary(lretsI)\ \
sd(lretsA)\ \
sd(lretsI)\ \
library(e1071)\ \
skewness(lretsA)\ \
skewness(lretsI)\ \
skewness(lretsA)\ \
kurtosis(lretsA)\ \
kurtosis(lretsI)\ \
cor(lretsA, lretsI)\ \
plot.ts(lretsA)\ \
plot.ts(lretsI)\ \
par(mfrow=c(1, 2))\ \
hist(lretA, nclass=40, freq=FALSE, main='Return histogram')\ \
hist(lretsA, nclass=40, freq=FALSE, main='Return histogram')\ \
curve(dnorm(x, mean=m, sd=s), from= -0.3, to=0.2, add=TRUE, col="red")\ \
plot(density(lretsA), main= 'Return empirical distribution')\ \
curve(dnorm(x, mean=m, sd=s), from=-0.3, to=0.2, add=TRUE, col="red")\ \
par(mfrow=c(1,1))\ \
hist(lretsI, nclass=40, freq=FALSE, main='Return histogram')\ \
m1=mean(lretsI)\ \
s1=sd(lretsI)\ \
s1\ \
hist(lretsI, nclass=40, freq=FALSE, main='Return histogram')\ \
curve(dnorm(x, mean=m1, sd=s1), from= -0.3, to=0.2, add=TRUE, col="red")\ \
plot(density(lretsI), main= 'Return empirical distribution')\ \
curve(dnorm(x, mean=m1, sd=s1), from=-0.3, to=0.2, add=TRUE, col="red")\ \
par(mfrow=c(1,1))\ \
library(normtest)\ \
jb.norm.test(lretsA, nrepl=2000)\ \
jb.norm.test(lretsI, nrepl=2000)\ \

sqlretsA=lretsA*lretsA\ \
sqlretsI=lretsI*lretsI\ \
library(LSTS)\ \
Box.Ljung.Test(sqlretsA, lag = 8, main = NULL)\ \
ts.diag(sqlretsA)\ \

Box.test(sqlretsA,type="Ljung",lag=8,fitdf=0)\ \
```



```

Box.test(sqLretsI,type="Ljung",lag=8,fitdf=0)\

ar(diff(lretsA, differences=2))\

library(aTSA)\

adf.test(diff(lretsA), 28)\
ar(x = diff(lretsI, differences = 2))\

adf.test(diff(lretsI), 28)\

library(forecast)\
acf(lretsA, lag=8)\
acf(lretsI, lag=8)\
pacf(lretsA, lag=8)\
pacf(lretsI, lag=8)\

acf(sqLretsA, lag=8)\
acf(sqLretsI, lag=8)\
pacf(sqLretsA, lag=8)\
pacf(sqLretsI, lag=8)\

library(rugarch)\

#Arma-Garch model for Asia and India\

#Arma-Garch for Asia\

model=ugarchspec (\
variance.model = list(model = "sGARCH", garchOrder = c(1, 1)),\
mean.model = list(armaOrder = c(3, 1)),\
distribution.model = "sstd")\
modelfit1=ugarchfit(model,data=lretsA,out.sample=0)\

modelfit1@fit$fitted.values\

Box.test(lretsA-modelfit1@fit$fitted.values,type="Ljung",lag=8,fitdf=0)\

model=ugarchspec (\
variance.model = list(model = "sGARCH", garchOrder = c(1, 1)),\
mean.model = list(armaOrder = c(1, 1)),\
distribution.model = "sstd")\
modelfit2=ugarchfit(model,data=lretsI,out.sample=0)\
Box.test(lretsI-modelfit2@fit$fitted.values,type="Ljung",lag=8,fitdf=0)\

residuals1= resid(lm(lretsA~modelfit1@fit$fitted.values))\
residuals2= resid(lm(lretsI~modelfit2@fit$fitted.values))\

```

```

#Residuals transformed to uniform student-t\\

u1=pt(residuals1, 1, ncp=0, lower.tail = TRUE, log.p = FALSE)\\
u2=pt(residuals2, 1, ncp=0, lower.tail = TRUE, log.p = FALSE)\\

#{Test to choose a copula\\
library(VineCopula)\\
library(copula)\\
u <- pobs(as.matrix(cbind(u1,u2)))[,1]\\
v <- pobs(as.matrix(cbind(u1,u2))[,2])\\

#{Fit copula\\
t.cop <- tCopula(dim=2)\\
set.seed(500)\\
m <- pobs(as.matrix(cbind(u1,u2)))\\
fit <- fitCopula(t.cop,m,method="ml")\\
coef(fit)\\
selectedCopula <- BiCopSelect(u,v,familyset=NA)\\
selectedCopula}\\

#{fit copula\\
t.cop <- tCopula(dim=2)\\
set.seed(500)\\
m <- pobs(as.matrix(cbind(u1,u2)))\\
fit <- fitCopula(t.cop,m,method="ml")\\
coef(fit)\\
rho <- coef(fit)[1]\\
df <- coef(fit)[2]\\
persp(tCopula(dim=2,rho,df=df),dCopula)\\

u <- rCopula(3965,tCopula(dim=2,rho,df=df))\\
plot(u[,1],u[,2],pch='.',col='blue')\\
cor(u,method='spearman')}\\

v1= rstandard(lm(lretsA~modelfit1@fit$fitted.values))\\
v2= rstandard(lm(lretsI~modelfit2@fit$fitted.values))\\

#Parameter estimation, mle method\\
#Clayton copula\\
est.mleV1 <- BiCopEst(v1, v2, family = 3, method = "mle")\\
summary(est.mleV1)\\

#Gumbel copula\\
est.mleV2 <- BiCopEst(v1, v2, family = 4, method = "mle")\\
summary(est.mleV2)\\

#Frank Copula\\
est.mleV3 <- BiCopEst(u1, u2, family = 5, method = "mle")\\
summary(est.mleV3)\\

#CopulaFit\\
C = matrix( c(v1, v2), nrow=2489, ncol=2)\\

```

```
cop_model <- gumbelCopula(1.81, dim=2)\\  
fit <- fitCopula (cop_model, C, method ="ml")\  
coef(fit)\\  
  
#for 3D\  
persp(gumbelCopula(1.81, dim=2),dCopula)\\  
  
#for Scatter plot\  
p = rCopula(3000,gumbelCopula(coef(fit),dim=2))\  
plot(p)\
```