Identifying Dependence Structure among Equities in Indian Markets using Copulas

Grover, Vaibhav

27 August 2015

Identifying Dependence Structure among Equities in Indian Markets using Copulas

Vaibhav Grover
vaibhavgrover25@gmail.com

Abstract

In this study we have examined that assets returns in Indian markets do not follow an elliptical dependence structure; asymmetric tail dependence can be observed among asset returns particularly when the assets exhibit downside returns in a bearish market. We have used Elliptical, Archimedean and Canonical Vine copulas to model such dependence structure in large portfolios. Using certain goodness-of-fit tests we find that Archimedean copulas are insufficient to model the dependence among assets in a large portfolio. We have also compared copula models using an out-of-sample Value-at-Risk (VaR) calculation and comparing results to the historical data. It is observed that the Canonical Vine copulas consistently capture the variation in weekly and daily VaR values.

Keywords: *copula, vine copulas, Value-at-Risk*

1. Introduction

Modern portfolio theory assumes that equity returns follow elliptical dependence. But, there is an increase in correlation of equity returns at the time of bearish markets which has been shown by many authors [Erb, Harvey, and Viskanta (1994), Longin and Solnik (2001), Ang and Bekaert (2002), Ang and Chen (2002), Campbell, Koedijk, and Kofman (2002), and Bae, Karolyi, and Stulz (2003)]. This increased correlations while equity returns are on a downside is known as the lower tail dependence and is in violation to assuming that the returns follow an elliptical distribution (Markowitz, 1952). Using elliptical (or normal) dependence and ignoring such correlations by the investors in forecasting models could lead to huge losses, whereas, inclusion of this asymmetric dependence could lead to significant gains or at least would reduce portfolio risk. In this study we have incorporated this tail dependencies for Indian markets using copulas.

Copulas, loosely speaking, defines the dependencies among the set of financial variables. The multivariate distribution of these variables can be fully specified using the marginal distributions of individual variables and by their copula. Modelling the marginal and the copula separately provides more flexibility. Different copula models like Archimedean copulas and Gaussian copula have been used to model such dependencies (Patton, 2004). But, the study usually has been limited to 3-4 variables. We have extended the study to a portfolio comprised of 8 National Stock Exchange (NSE) industry indices. We have also used more advanced Canonical Vine Copulas (introduced by Aas et al, 2009) to model comparatively large portfolios.

Our study is motivated from work of Low, Alcock, Faff and Brailsford (2013) in US markets. We have explored the answers to the following questions in Indian markets: whether asymmetric dependence or lower tail dependence is exhibited by Indian Equity Markets? If yes, then whether existing copula models could be used to model this dependence and forecast Value-at-Risk (VaR) in future? Which copula is best suited to model the asset returns of a large portfolio? Or, one would need a mixture of copulas? Is a single parameter Archimedean copula enough to capture dependencies of an 8 assets portfolio?

The rest of the paper is divided as follows: Section 2 is a brief description of basic copula theory and copula models; Section 3 is an exploratory analysis of the data where we have shown that the data does not follow a normal dependence structure and shown the evidence of points present in tails of correlation plots; Section 4 describes the various steps we have followed to fit various copulas to data and modelling the individual CDF function of each asset; Section 5 depicts how we have used the fitted distributions to calculate weekly and daily Value-at-Risk and Section 6 is the conclusion.

2. Copula Theory

According to Sklar (1973), for n ≥ 2, let G be an n-dimensional distribution function with 1-margins F_1 ..., F_n. Then there exists an n-copula C such that,

$$G(x_1, \dots, x_n) = C[F_1(x_1), \dots, F_n(x_n)]$$

for all tuples (x_1 ,...,x_n) in E^n. So, any cumulative distribution function can be broken down into the distribution function of its components or marginals and the dependence structure between these components, known as copula. Hence, this approach provides us with the power to choose marginal distribution and then independently model the dependence structure between the components providing more flexibility to the model. According to Sklar, a joint distribution can be written in terms of marginal distributions and copula distribution function,

$$f(x_1, \dots \dots \dots, x_n) = f_1(x_1) \times f_2(x_2) \times \dots \dots \dots \times f_n(x_n) \times c[F_1(x_1), \dots, F_n(x_n)]$$

Where, $c[F_1(x_1), \dots, F_n(x_n)] = \frac{\partial^n C[F_1(x_1),\dots,F_n(x_n)]}{\partial x_1 \partial x_2 \dots\dots\dots \partial x_n}$

2.1. Elliptical Copulas

The most common elliptical copulas are Gaussian and Student-t which are derived from multivariate normal and student-t distributions. The advantage of elliptical copulas is that one can specify different level of correlation between the marginals but the disadvantage being that they have radial symmetry. For a given correlation matrix $R \in \Re^{d \times d}$, the Gaussian copula with parameter matrix R can be written as,

$$C_R^{Gauss}(\boldsymbol{u}) = F_R(F^{-1}(u_1), \dots \dots, F^{-1}(u_d))$$

where F^{-1} is the inverse cumulative distribution function of a standard normal and F_R is the joint distribution of a multivariate normal distribution with mean vector zero and covariance matrix equal to correlation matrix R. The density can be written as,

$$c_R^{Gauss}(\boldsymbol{u}) = \frac{1}{\sqrt{detR}} \exp[-\frac{1}{2}(F^{-1}(u_1) \quad \dots \quad F^{-1}(u_d)).(R^{-1} - I).\begin{pmatrix} F^{-1}(u_1) \\ \vdots \\ F^{-1}(u_d) \end{pmatrix}]$$

Similarly, a student-t copula with univariate student–t distribution as marginals can be written as

$$c_{v,P}^t(\boldsymbol{u}) = \frac{f_{v,P}(t_v^{-1}(u_1), \dots, t_v^{-1}(u_d))}{\prod_{i=1}^d f_v(t_v^{-1}(u_i))}, \qquad \boldsymbol{u} \in (0,1)^d$$

where $f_{v,P}$ is the joint distribution of a $t_d(v, \mathbf{0}, P)$- distributed random vector and f_v is the density of univariate standard t-distribution with v degrees of freedom.

2.2. Archimedean Copulas

Archimedean copulas allow modeling dependence in arbitrarily high dimensions with only one parameter, governing the strength of dependence. Most of the Archimedean copulas admit an explicit formula. A copula C is Archimedean if it admits the following representation,

$$C(u_1, \dots \dots, u_d; \theta) = F(F^{-1}(u_1;\theta) + \cdots + F^{-1}(u_d;\theta);\theta)$$

F is called the generator function and satisfies:

- $F: [0,\infty) \to [0,1]$ with $F(0) = 1$ and $\lim_{x\to\infty} F(x) = 0$
- F is a continuous
- F is strictly decreasing on $[0, F^{-1}(0)]$
- F^{-1} is pseudo inverse defined by $F^{-1}(x) = \inf\{u : F(u) \le x\}$

Famous Archimedean copulas and their generator functions are listed in Table 1.

Table 1
Archimedean Copulas: Generator functions and parameter range

Copula name	Generator Function $F(t)$	Generator function inverse $F^{-1}(t)$	Parameter Range (θ)
Clayton	$(1+\theta t)^{-1/\theta}$	$\frac{1}{\theta}(t^{-\theta}-1)$	$\theta \in [-1,\infty)\backslash\{0\}$
Ali-Mikhail-Haq	$\frac{1-\theta}{\exp(t)-\theta}$	$log(\frac{1-\theta(1-t)}{t})$	$\theta \in [-1,1)$
Gumbel	$(-\log(t))^{\theta}$	$\exp(-t^{1/\theta})$	$\theta \in [-1,\infty)$
Frank	$\frac{1}{\theta}\log(1+\exp(-t)\,(\exp(-\theta)-1))$	$-\log(\frac{\exp(-\theta t)-1}{\exp(-\theta)-1})$	$\theta \in R\backslash\{0\}$
Joe	$1-(1-\exp(-t))^{1/\theta}$	$-\log(1-(1-t)^{\theta}$	$\theta \in [-1,\infty)$
Independence	$\exp(-t)$	$-\log(t)$	

2.3. Canonical Vine Copulas

Vines introduced by Aas et al. (2009), are a graphical representation to specify a pair copula constructions (PCCs). First we explain pair copula construction by using an example $X = (x_1, x_2, x_3) \sim F$ with marginal distribution functions F_1, F_3 and F_3 and corresponding densities. So we can write,

$$f(x_1,x_2,x_3) = f(x_1)f(x_2|x_1)f(x_3|x_1,x_2).$$

By Sklar's theorem, we know that

$$f(x_2|x_1) = \frac{f(x_1,x_2)}{f_1(x_1)} = \frac{c_{1,2}(F_1(x_1),F_2(x_2))f_1(x_1)f_2(x_2)}{f_1(x_1)} = c_{1,2}(F_1(x_1),F_2(x_2))f_2(x_2)$$

And

$$f(x_3|x_1,x_2) = \frac{f(x_2,x_3|x_1)}{f_1(x_2|x_1)} = \frac{c_{2,3|1}\big(F(x_2|x_1),F(x_3|x_1)\big)f(x_2|x_1)f(x_3|x_1)}{f_1(x_1)}$$

$$= c_{1,2}\big(F_1(x_1), F_2(x_2)\big)f_2(x_2)$$

$$= c_{2,3|1}\big(F(x_2|x_1), F(x_3|x_1)\big)f(x_3|x_1)$$

$$= c_{2,3|1}\big(F(x_2|x_1), F(x_3|x_1)\big)\, c_{1,3}\big(F_1(x_1), F_3(x_3)\big)f_3(x_3).$$

Therefore, it is possible to represent the 3-dimensional joint distribution using bivariate copulas $C_{1,2}, C_{1,3}$ and $C_{2,3|1}$ which are known as pair copulas. It is possible to choose these pair copulas independently of each other which provides us with wide range of dependence structure. It is usually assumed that conditional copula $C_{2,3|1}$ is independent of conditioning variable X_1 to facilitate inference.

Since decomposition of $f(x_1, x_2, x_3)$ is not unique many possible PCCs are possible so for classification graphical models called vine were introduced. Vines arrange the $d(d-1)/2$ pair copulas of d-dimensional PCC in *d-1* linked trees (acyclic graphs). In the first C-Vine tree, bivariate copulas with respect to a root node is calculated for all the other variables. Then conditioned on root node of first tree, a second variable is chosen with which all pairwise dependencies are modelled in the second tree. So a root node is chosen in all trees and dependencies are modelled with respect to this node conditioned on all previous root nodes. So, C-vine density with *1,…,d* root nodes is written as,

$$f_{12\ldots\ldots d}(x) = \prod_{k=1}^{d} f_k \prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{j,j+1|1,\ldots,j-1}(F\big(x_j|x_1, \ldots., x_{j-1}\big), F\big(x_{j+1}|x_1, \ldots., x_{j-1}\big))$$

where, $c_{j,j+1|1,\ldots,j-1}$ represent the bivariate conditional copulas and f_k depicts the marginal densities. The model we have explained is Canonical Vine (C-Vine) model which we will use for data analysis further.

3. Data

The data we have chosen consists daily prices for 8 Indian market indices, namely: Automotive, Bank, Energy, Finance, FMCG (Fast-moving consumer goods), IT, Metal and Pharma. Indian market has 11 sector indices but because of lack of data for the others, 8 of them have been taken for this study. The data dates from January 2005 to December 2014 (collected from http://www.nseindia.com/), which gives a total of 2480 observations of daily closing prices.

For the analysis, we have considered indices as assets which provides us with the following advantages:

- Considering particular stocks would make the study susceptible to risks specific to an asset and that have no correlation to market risks, known as idiosyncratic risks.
- Study would be applicable to the whole market rather than a small section of market.

The daily returns for each index are then calculated. Table 2 shows the mean, standard deviation, minimum, maximum, skewness and kurtosis of daily returns for each industry index.

Table 2:
Descriptive Statistics of Index Returns

	Mean	Standard Deviation	Skewness	Kurtosis	Minimum	Maximum
Automotive	0.6523	4.4616	-0.3551	5.5699	-22.7668	18.0677

Bank	0.6424	5.9248	0.3230	5.1241	-20.5557	27.3827
Energy	0.3541	5.2803	-0.2830	12.3166	-31.9490	32.6477
Finance	0.6593	5.6222	0.2510	5.0620	-18.1852	25.4256
FMCG	0.6259	3.6835	-0.4523	5.2961	-16.6246	13.4282
IT	0.4727	4.4608	-0.0828	4.8264	-14.5032	19.5607
Metal	0.4648	6.7575	-0.0021	6.3871	-31.5294	32.2389
Pharma	0.5148	3.5493	-0.9613	7.1154	-17.2002	11.4255

Table shows the descriptive analysis of daily returns of 8 Indian industry indices. Mean, minimum and maximum are shown as percentages.

All indices except Bank and Finance are negatively skewed. Excess kurtosis is shown by every index return. The observations are tested for normality by Jarque-Bera and Kolmogrov-Smirnov tests against 5% significance level. All indices reject the null hypothesis (Null Hypothesis: Sample belongs to a normal distribution) for both tests of normality at 5% as well as 1% significance level. The value of test statistics are summarized in Table 3.

Table 3:
KS and JB Test Statistics (5% Significance level)

	Jarque-Bera Test Statistic	Kolmogrov-Smirnov Test Statistic
Automotive	104.9	0.3666
Bank	72.7	0.3605
Energy	1285.0	0.3238
Finance	66.4	0.3713
FMCG	89.8	0.3158
IT	49.6	0.3496
Metal	169.2	0.3449
Pharma	304.3	0.3365

Table 4 & Figure 1 depict the pairwise Pearson's correlation coefficients of historical index returns over the whole set of observations.

If we have a closer look at the histograms in the above figure they also hint towards the observations not following a normal distribution. Also, the dependence between returns of any two assets does not follow a normal distribution. This is coined as 'fan-shaped' behavior and indicates the presence of tail dependence. In some cases, the tail is extended more towards the lower side (negative return values) as compared to the upper side (positive return values) which shows that the tail is asymmetric, i.e., extended more towards the region of negative index returns. This shows that as in US markets, the correlations between industry indices is higher in bearish markets as compared to bullish markets (supports earlier studies).

4 Multivariate Joint Distribution

The first step is to model the joint distribution for the assets and for that we followed the Inference for margins (IFM) proposed by Joe and Xu (1996). The process is a 2-step approach that includes:

1. Marginal modelling
2. Copula modelling

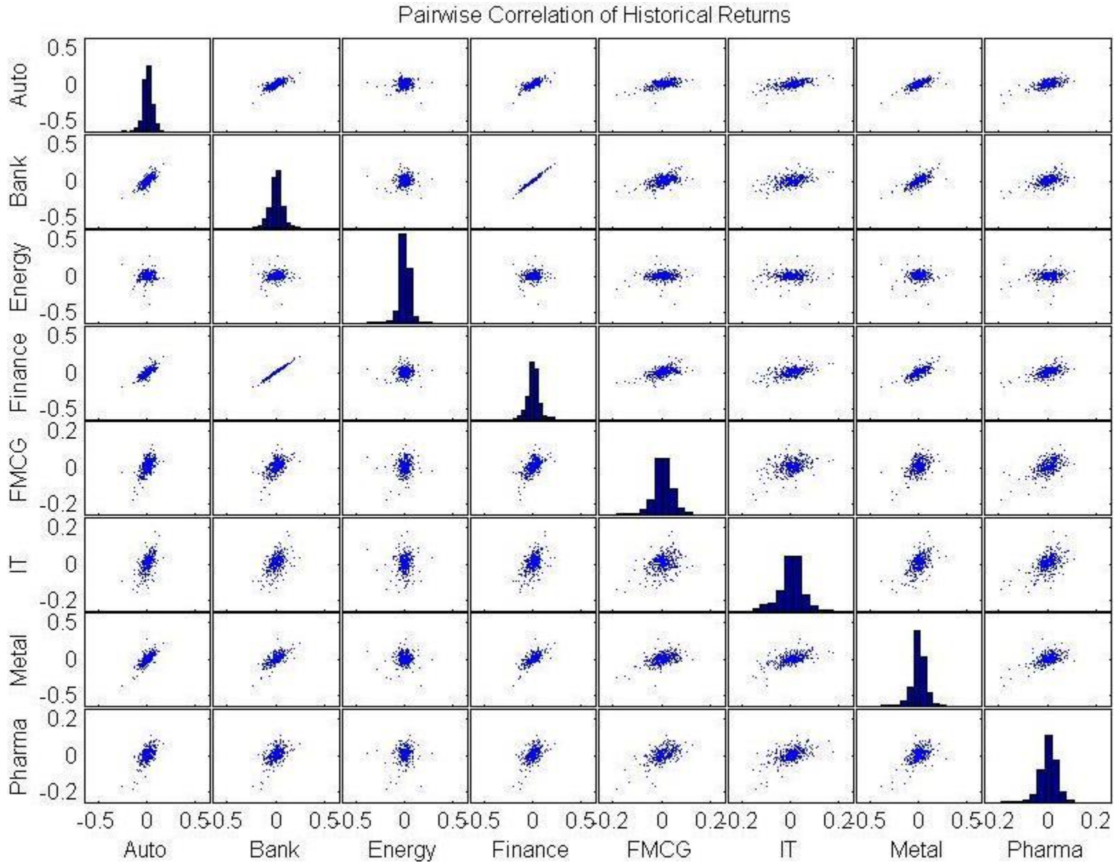


Figure 1:
The figure shows the histograms for each historical index return and the pairwise correlation plots.

Table 4:
Pairwise correlation of historical returns

	Automotive	bank	Energy	Finance	FMCG	IT	Metal	Pharma
Automotive	1	0.7445	0.0475	0.7661	0.5852	0.5360	0.7198	0.6119
Bank		1	0.0481	0.9870	0.5307	0.4654	0.6914	0.5165
Energy			1	0.0573	0.0982	0.0332	0.0130	0.1307
Finance				1	0.5672	0.5024	0.7157	0.5352
FMCG					1	0.3673	0.4648	0.5393
IT						1	0.5516	0.5398
Metal							1	0.5180
Pharma								1

The IFM method estimates the marginal parameters in the 1st step and then estimated the parameters of the copula in the 2nd step. But in our approach instead of using the Maximum likelihood estimation we have used a different approach for fitting marginal and copula parameter estimation which is discussed in the next two upcoming sections.

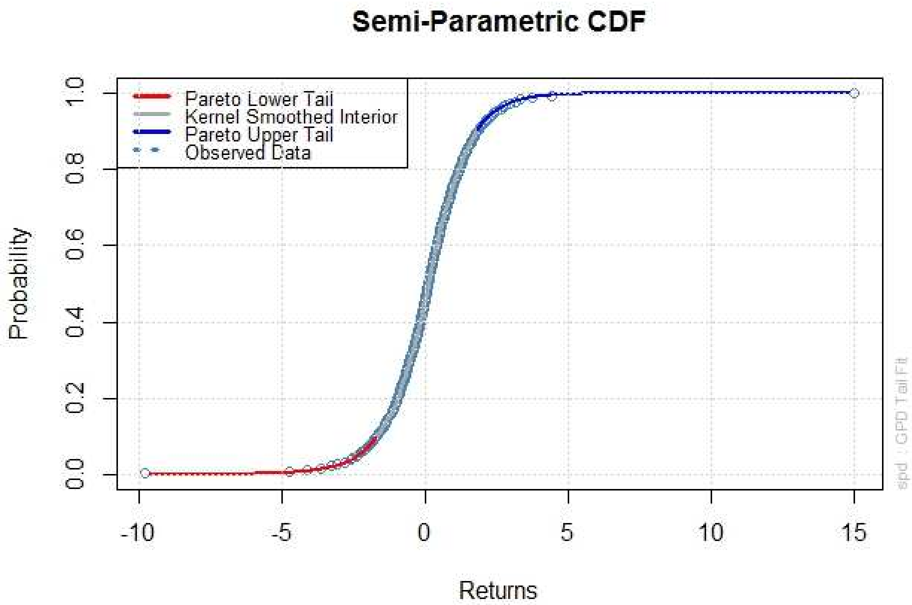
4.1 Marginal Modelling

We first have to model the marginal distributions. We have modelled the empirical cumulative distribution of each marginal using a Semi-Parametric Distribution (SPD) function. The semi-parametric function has 3 parts, the lower tail, interior distribution and upper tail. The upper and lower tails each is 10% of the total number of observations that are fitted and is modelled using Generalized Pareto Distribution (GPD). The interior part is a smoothened version of step function, i.e., empirical cumulative distribution with the help of a Gaussian kernel SPD.

Using GPD gives us the benefit of being able to model the tail in-spite of having less number of observations in tails. The resulting distribution also allows the extrapolation in each tail which allows estimation of returns outside the historical record. Cumulative distribution function for 'Automotive' index returns is shown in Figure-2. Its SPD is given by,

```
Piecewise distribution with 3 segments:
         -Inf < x < -1.69 (0 < p < 0.1)   : lower tail, GPD(0.116, 0.894)
      -1.69 < x < 1.83   (0.1 < p < 0.9)  : interpolated kernel smooth cdf
        1.83 < x < Inf    (0.9 < p < 1)   : upper tail, GPD(0.112, 0.803)
```

Figure 2:
Auto index return cdf



4.2 Copula Estimation

The following copulas were considered to model the dependence structure of 8 index returns: Elliptical copulas: Normal and Student-t; Archimedean copulas: Gumbel, Clayton, Joe and Frank. All of these mentioned copulas are 1-parameter or 2-parameter specification of dependence structure. 'Inverse Kendall's Tau' approach was applied to calculate the parameter values for the whole dataset of 2479 return observations (Genest and NeSlehova 2011).

After that, goodness-of-fit test was applied to all fits calculating the S_n statistic value (Equation (2) in Genest, Remillard and Beaudoin (2009)). The statistic basically provides he distance between the fitted copulas and the empirical copula attained using the data. The results are combined in the following Table 5.

4.3 Canonical Vine Copulas Estimation

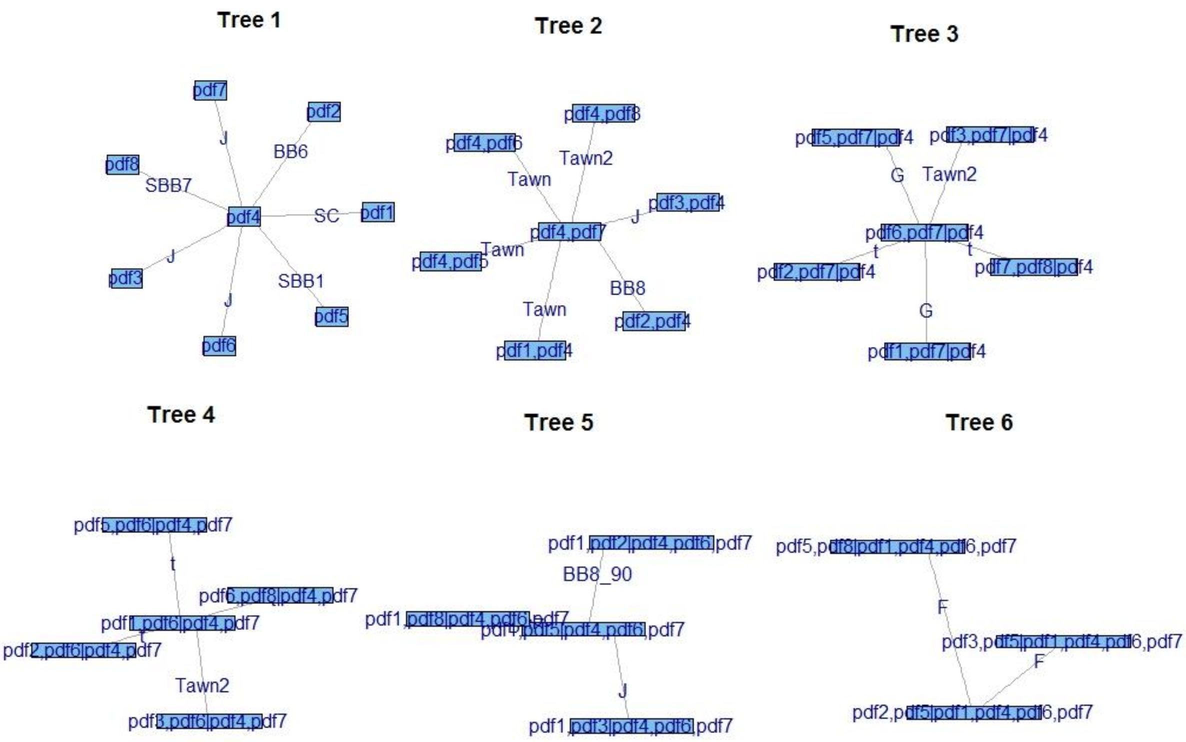
As mentioned earlier, the Archimedean and Elliptical copula families are defined by 1 or 2 parameters. We calculated the p-values of goodness-of-fit test for Archimedean and Elliptical copulas and the insignificant p-values show that for capturing a dependence structure of 8 assets such copulas are not accurate. Because of this reason we propose to model the dependence structures using Canonical Vine Copulas (CVC).

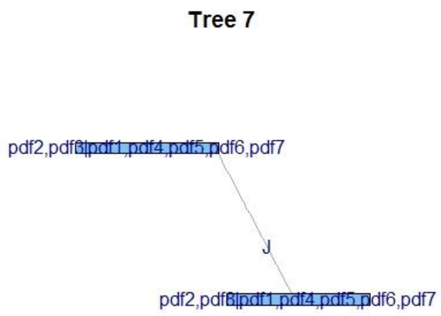
To check how good CVCs model the dependence structure, we have fitted the whole set of observations to the CVC keeping the marginal which has the maximum correlation with the others at the root. For each bivariate pair, a copula out of 39 available copulas (list provided at the end) in the package 'VineCopula' in R was selected on the basis of AIC (Akaike Information Criterion). The generated tree structures are shown in the figure below.

Table 5:
Goodness-of-fit statistic

Copula	Parameter Value	S_n
Normal	0.301	0.808
Student-t	0.301	0.851
Gumbel	1.33	1.26
Clayton	1.44	0.739
Frank	2.15	1.08
Joe	1.61	1.9

Figure 3:
CVC tree structures

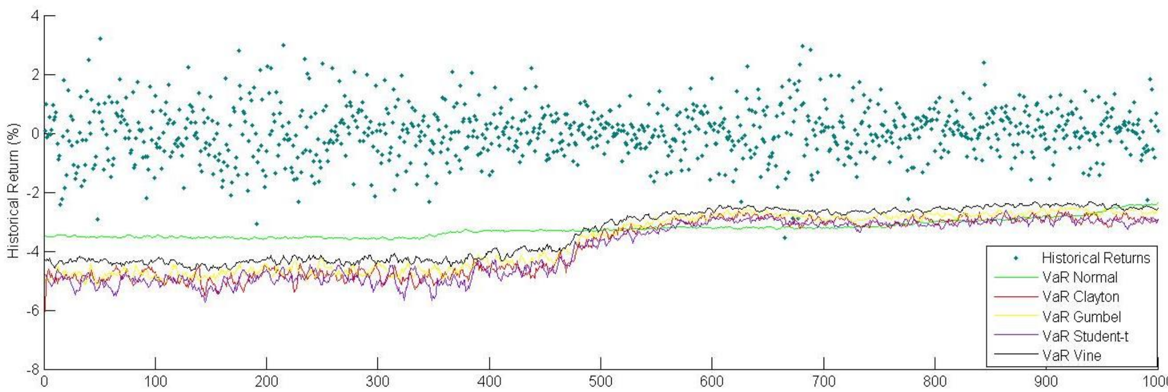




5. Empirical Results

After fitting data to various copulas we have simulated using the fitted distributions to calculate Value-at-Risk (VaR). For this we have used a rolling window approach taking first 1479 daily return observations as training data and fitting a copula for the same data. Then we calculate the VaR for the next day using simulated from the fitted copula. After that, removing the first value of daily return and including 1500th value we fitted a new copula and then calculated VaR for the next day. This provided us with 1000 daily VaR values. The copulas used for the simulation are CVC's, Clayton, Gumbel, Student-t and Multivariate Normal Distribution. Figure 4 compares the historical returns with the smoothened VaR plot obtained from different copulas. Because of high fluctuation in VaR values, there was an overlap of plot lines, so to show a distinction, the plot has been smoothened using a moving average filter. The portfolio we have chosen is an equal weighted portfolio of 8 sector indices from Indian markets

Figure 4.
Daily Value-at-Risk



As we can see from figure 4, the plot of VaR (with time on x-axis) calculated using Canonical Vine Copulas (except multivariate normal distribution) is closest to the historical returns of the 8 asset equal weighted portfolio under consideration. In case of single or two parameter copulas, Gumbel copula performs better than Student-t copula and Clayton copula, whereas, the plot lines of the both copulas are almost overlapping. This shows that a single or two parameter copula is not able to capture the dependence structure of 8 assets together.

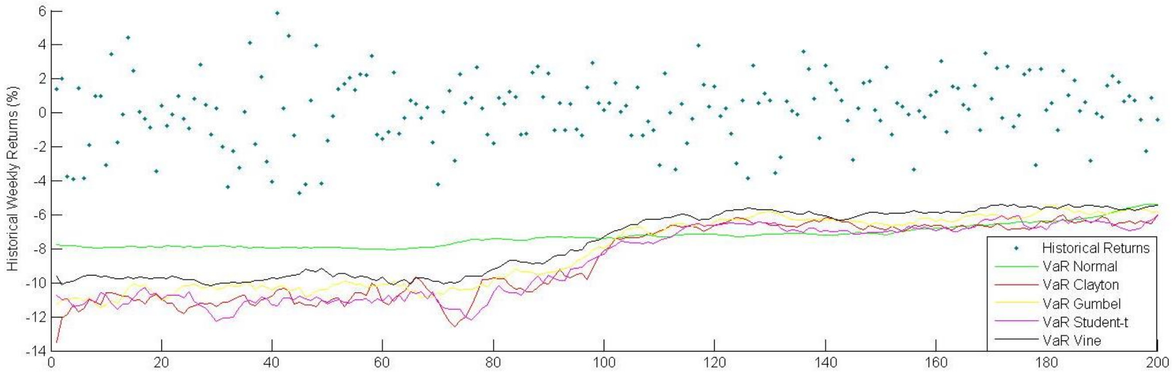
Comparing the performance of copulas to the multivariate normal distribution, it is can be seen that is almost constant for normal distribution. Hence, if we assume that the assets follow a normal distribution, we would be unable to capture the tails of asset returns.

Same observations can be made when instead of calculating daily VaR, we used the fitted distributions to calculate the weekly VaR. The daily VaR values can easily be converted to the weekly values by the following formula,

$$VaR_{weekly} = \sqrt{5} \times VaR_{daily}$$

We considered a week of 5 working days so instead of 1000 observations we had 200 observations of VaR values. The comparison of weekly VaR values to weekly historical returns in shown in Figure 5.

Figure 5
Weekly Value-at-Risk



6. Conclusion

In this study, firstly we have inspected the return distribution of 8 Indian market sector indices. We have found asymmetric in return distributions using exploratory analysis and using 2 statistical tests showed that it is not possible to represent returns using normal distribution. After this, we have tried to model return distributions using Archimedean and Elliptical copulas that are defined using single or double parameters. The goodness of fit tests done on the fitted distributions suggest that such copulas are not sufficient to model dependence structures of portfolio of 8 assets. So, we have considered Canonical Vine Copulas to model the dependence of return distributions.

To compare the models of returns, we have calculated weekly and daily Value-at-Risk values of a portfolio of 8 assets having equal weights by drawing random samples from each distribution. Comparing the VaR values to the historical returns of the portfolio we observed that copulas are able to capture the asymmetric dependence structures better than a multivariate normal distribution. Out of the copulas tested, CVC's provide VaR values closest to the historical returns proving that for a large portfolio, such as that of 8 assets, single copulas are not enough to model the dependence.

References

[1] Beaudoin, D., Genest, C. and Remillard, B. (2009). "Goodness-of-fit tests for copulas: A review and a power study", *Insurance: Mathematics and Economics 44(2), Pages 199–213*

[2] Alcock, J., Brailsford, T., Faff, R., and Low, R. (2013). "Canonical vine copulas in the context of modern portfolio management: Are they worth it?" *Journal of Banking & Finance 37 (2013), Pages 3085–3099*

[3] Erb, C., Harvey, C. and Viskanta, T. (1994), "Forecasting International Equity Correlations", *Financial Analysts Journal 50(6)*

[4] Longin, F. and Solnik, B. (2001), "Extreme Correlation of International Equity Market", *The Journal of Finance* Vol. 56, No. 2 (Apr., 2001), pp. 649-676

[5] Ang A. and Bekaert, G. (2002), "International Asset Allocation with Regime Shifts", *Review of Financial Studies Volume 15(4)*, Pg. 1137-1187

[6] Ang, A. and Chen, J. (2002), "Asymmetric correlations of equity portfolios"**,** *Journal of Financial Economics Volume 63, Issue 3*, Pages 443–494

[7]Campbell, R. Koedijk, K. and Kofman, P. (2002), "Increased Correlation in Bear Markets", *Financial Analysts Journal 58(1)*

[8] Bae, K., Karolyi, G., and Stulz, R. (2003), "A New Approach to Measuring Financial Contagion"*, Financial Analysts Journal 16(3)*, Pg. 717-763

[9] Markowitz, H. (1952), "Portfolio Selection", *The Journal of Finance 7(1)*, Pg. 77-91

[10] Patton, A. (2004), On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation *Volume 2(1),* Pg 130-168

[11] Aas, K., Bakken, H., Czado, C. and Frigessi, A. (2009), "Pair-copula constructions of multiple dependence, Insurance: Mathematics and Economics*" 44(2)*, Pg.182-198

[12] Joe, H. and Xu, J. (1996). "The Estimation Method of Inference Functions for Margins for Multivariate Models." *Technical Report 166, Department of Statistics, University of British Columbia*.

[13] Yan, J. (2007), "Enjoy the Joy of Copulas: With a Package copula", *Journal of Statistical Software Volume 21(4)*

[14] Brechmann, E. and Schepsmeier, U. (2013), "Modeling Dependence with C- and D-Vine Copulas: The R Package CDVine", *Journal of Statistical Software Volume 52(3)*

[15] Genest, C. and NeSlehova, J. (2011), "Estimators based on Kendall's tau in multivariate copula model, *Australian & AMP New Zeland Journal of Statistics*