Analysis of Multivariate Financial Data Using Different Copulas

Annemaria Mathews & Harsh Roghelia

April 20, 2018

Abstract

This paper reviews the various copula models on multidimensional asset returns. Asset returns are very often correlated with each other; however, it can be difficult to observe this dependencies. Transforming the marginal distribution of the asset returns using copulas removes the effect of the joint distribution and allows one to model the dependencies. There are several copula classes availabe and each has varying degrees of success in modeling financial data. We compare the performance of elliptical and Archimedean copula classes during periods of average volatility and periods of high volatiltity in the market. We also compare the performance of these copulas when they are fit based on variable dependencies. We test the copula model by using a variant of the Cramér-von Mises (CvM) test and test the overall financial model using VaR simulations. We conclude that further investigation will be needed to support our findings.

Introduction

Large amounts of financial data are consumed by financial and investment firms to build models that give insights to market conditions. A combination of technical, fundamental, and statistical analysis is used to determine the best investment strategy that balances risk and return. Analysts strive to build models that are close to reality, however, these models are at best a rough estimate of true market conditions because assumptions need to be made to fit the data and to keep the program computationally feasible. On the other hand, making these assumptions can lead to immense loss, as witnessed in the financial crisis of 2008. The assumption that stock returns are normally distributed was widespread and deeply entrenched in financial models. In truth, while stock returns are normal during periods of stability, empirical evidence shows they can be skewed and heavy-tailed during periods of

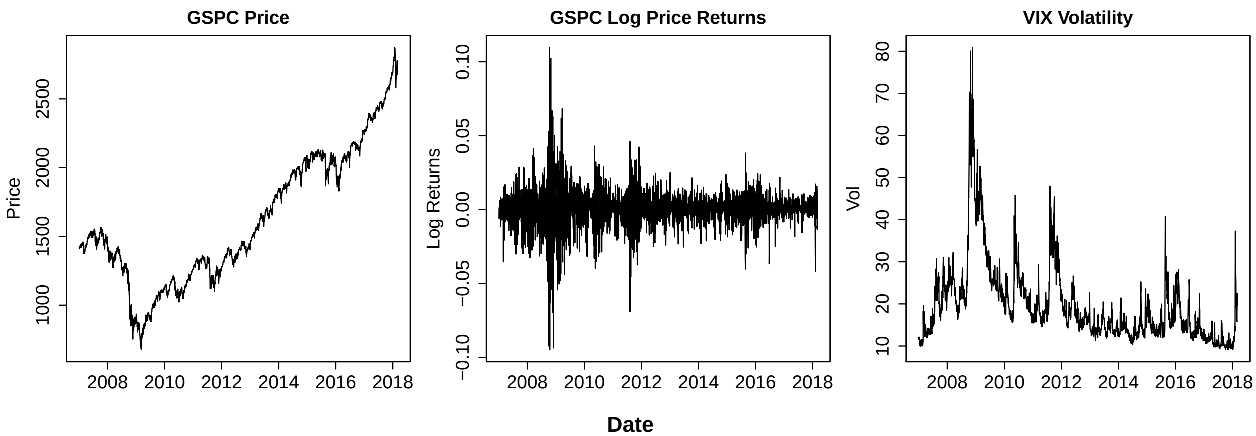

Figure 1: The GSPC Price, GSPC Log Price Returns and VIX Volatility plots

high volatility. Additionally, stocks have different correlations with each other during periods of stability but tend to move together during periods of volatility (Palaro & Hotta, 2006). In this paper, we attempt to alleviate this problem by comparing GARCH models that use various copula methods.

Using daily prices of 12 stocks from the S&P 500 between 2012 and 2017, we compare GARCH models that use elliptical (normal and t) and Archimedean (Gumbel and Clayton) copulas. We evaluate the models by conducting an empirical copula goodness of fit test based on the CvM criterion described in Genest et al. (2009). The performance of the overall model (GARCH+copula model) is tested by comparing out-of-sample return forecasts with observed returns based on the methodology described in Stulajter (2003). We picked the period between 2012 to 2017 because the relatively long period smooths out abnormal returns and market movements. Therefore we will call this the period of normal volatility.

We also fitted the four copula classes to the period between 2007 and 2010 to test their performance during periods of high volatility (see Figure 1) the Furthermore, copula models can be difficult to fit when dealing with high dimensional data (Hofert 2012). We check if it is possible to building a high dimensional model by fitting a different copula to each sector. Here, we assume that assets within a sector move together and thereby have the same dependencies.

We suspected the Elliptical copulas would do better in periods of stability because it takes into account different pairwise correlation of stocks and Archimedean would do better during periods of volatility when stocks tend to have the same correlation since Archimedean copulas assume correlations are equal. However, upon further investigation, we found that the CvM is not completely conclusive and needs to be supplemented with more goodness of fit tests. Results from the VaR simulation show that t and Gumbel distribution perform well during peiods of average and high volatility. This did not fit with our intial hypothesis and assumptions. In the future, we would like to test the model with more data to find more evidence for our conclusions.

Definition of Copulas

Copulas were first introduced by Sklar in 1959, where he stated that any multivariate joint distribution $\mathbf{X}$ can be factored into two parts, the marginal distribution functions F_i which describes the univariate features of $\mathbf{X}$ and the copula function, C, which describes the dependencies of the distribution $\mathbf{X}$ (Nelsen 2003).

Mathematical Definiton: Let $\mathbf{X} = (X_1, ..., X_n)$ be a random vector. Let F denote the distributional function of $\mathbf{X}$ with marginal distribution functions F_i, such that $X_i \sim F_i$, 1 $\leq$ i $\leq$ n. By applying *Probability Integral Transform*, we get uniformly distributed marginals,

$$U_i = F_i(X_i), \quad 1 \leq i \leq n$$

. The copula function, C, of $\mathbf{X}$, is the joint cumulative distribution function of U_i,

$$C(u_1, u_2, ..., u_n) = Pr[U_1 \leq u_1, ..., U_n \leq u_n]$$

The inverse of these steps also holds true. Given U_i from the copula distribution, the pseudo-random samples can be obtained as,

$$(X_1, X_2, ..., X_n) = (F_1^{-1}(U_1), F_2^{-1}(U_2), ..., F_n^{-1}(U_n))$$

where F_i^{-1} denotes the generalized inverse of F_i, i.e. the quantile transform which is defined as,

$$F_i^{-1}(t) = inf\left\{x \in \mathrm{I\!R}^1; F_i(x) \geq t\right\}$$

Since the F_i's are assumed to be continuous, the copula function can be rewritten as,

$$\begin{aligned} C(u_1, u_2, ..., u_n) &= Pr[U_1 \leq u_1, ..., U_n \leq u_n] \\ &= Pr[X_1 \leq F_1^{-1}(u_1), ..., X_n \leq F_n^{-1}(u_n)] \end{aligned}$$

Hence, C is a copula of F where

$$F(x_1, ..., x_n) = P(X_1 \leq x_1), ...X_n \leq x_n)$$

Families of Copulas

There exists two popular methods for the derivation of copulas. The first is the inversion method which derives the copula from multivariate normal or student-t distributions. These copulas are commonly referred to as elliptical copulas. The other method uses a generator function to derive the copulas and these are known as Archimedean copulas.

Elliptical Copulas

The elliptical copulas are derived using the inversion method from the multivariate and univariate distribution functions which are denoted by Φ and ϕ respectively. The copula is defined as follows:

$$C(u_1, ..., u_2; \Sigma) = \Phi(\phi^{-1}(u_1), ..., \phi^{-1})$$

where Σ denotes the correlation matrix and $u_i = F_{x_i}(x_i)$ for $i = 1, .., n$; F_{x_i} denotes the cumulative distribution functions of the marginals (Stulajter 2009). In this report, we will analyze the normal and student-t copula functions. Hence, in the case of normal, Φ and ϕ is the multivariate and univariate standard normal, respectively. Similarly, for the case of student-t, $\Phi = \mathbf{t}_v$ and $\phi = t_v^{-1}$, where $\mathbf{t}_v$ and t_v are the multivariate and univariate Student t distributions with v degrees of freedom.

Archimedean Copulas

The Archimedean copulas are derived using the generator function Ψ, and is defined as follows:

$$C(u_1, ..., u_n) = \Psi^{-1}(\psi(u_1) + ... + \psi(u_n))$$

There are various Archimedean copulas, but we will analyze Clayton and Gumbel copulas. Their corresponding generator functions are (Stulajter 2009):

$$\begin{aligned} \Psi_{Clayton}(u) &= \alpha^{-1}(u^{-\alpha} - 1), \quad \alpha \neq 0 \\ \Psi_{Gumbel}(u) &= -(ln(u))^{\delta}, \qquad \delta \geq 1 \end{aligned}$$

The GARCH(1,1) model

Financial data is known to exhibit stochastic volatilities as it constantly shifts between stable and unstable price levels (Posedel1 2005); thus, it is nearly impossible to model this using a purely stationary multivariate normal. GARCH models, therefore, is more effective in modeling asset returns compared to ordinary least square analysis which assumes constant volatility (Engle 2001). In addition, the problem is simplified because unlike many other stochastic volatility models, GARCH has a closed form likelihood and, therefore, does not require a latent volatility to be integrated out.

Let $s_{ti} > 0$ denote the value of asset i on day t, and $y_{ti} \in \mathbb{R}$ denote its log returns, i.e.,

$$y_{ti} = log(s_{ti}) - log(s_{t-1,i})$$

Then the GARCH(1,1) model for return series of asset i is

$$y_{ti} = \mu_i + \varepsilon_{ti}\varepsilon_{ti} = \delta_{ti}x_{ti}\sigma^2_{ti} = \omega_i + \alpha_i\varepsilon^2_{t-1,i} + \beta_i\sigma^2_{t-1,i}x_{ti} \overset{iid}{\sim} f(x|\boldsymbol{\eta}_i)$$

Here, $f(x|\boldsymbol{\eta})$ is a noncentralized-t distribution as it allows for heavy-tailed returns and asymmetry, which is often the case in financial data.

Fitting GARCH(1,1) + Copula to Asset Returns

The four copulas (normal, student t, Clayton, and Gumbel) were fitted to log returns of assets in the S&P 500 for two different periods between 2007 and 2017. The dataset contained three assets each from the financial, technology, industrial, and health sectors (for 12 total assets). We assumed that assets within each sector have similar correlations. We used this assumption later on to build a model that fits separate copulas based on sectors.

GSPC and VIX volatility movements through the years show that asset returns between 2012 and 2017 capture varying volatility periods. Because it is a relatively long period, we assume any abnormal volatilities are smoothed out. We call this period the period of normal volatility. We fit the GARCH(1,1) model with skewed-t distribution to the log returns of each asset from the normal volatility period. Pseudo-observations were made of the standardized residuals obtained from the GARCH(1,1) model. This transforms the residuals to a $U(0,1)$ distribution and removes the effect of the assets' underlying distribution on their correlations (Nelsen 2003). Next, the corresponding elliptical and Archimedean copulas were fitted to the transformed residuals to obtain the optimal parameter estimates for each of the copulas. Our goal was to understand how various GARCH+copula models perform under normal market conditions. Performance of the models was measured by doing a CvM Goodness-of-Fit (GoF) test and comparing the models' simulated VaRs with observed returns from June 2017 to March 2018. In addition, we considered the possibility that elliptical and Archimedean copulas may perform differently when assets from each sector are fitted with a different copula. Thus, we determined if GARCH with various copulas result in a better fit than GARCH with a single copula.

Each GARCH+copula model was also fit to data from 2007 to 2010 (called period of high volatility) to understand the performance of these models under abnormal and volatile market conditions. There is evidence that assets tend to move together during periods of high volatility (Palaro & Hotta 2006) and this can affect the goodness of fit of each copula model. In addition to the CvM GoF test, the models were tested against actual returns from 2010 to 2011.

Goodness of Fit

In this paper, we conduct the GoF test by considering the empirical copula, i.e. pseudo-observations of the data which is summarized by

$$C_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(U_{i1} \leq u_1, \ldots, U_{id} \leq u_d)\mathbf{u} = (u_1, ..., u_d) \in [0,1]^d$$

It is considered a good estimator for the true underlying copula, C, of the data; therefore, we will use this as a benchmark to test $(H_0 : C \in \mathcal{C}_0)$ the estimated model (Genest et al. 2009). The empirical copula is used in a rank based version of the Cramer-von Mises test with the statistic,

$$S_n = \int_{[0,1]^d} n(C_n(\mathbf{u}) - C_{\theta_n}(\mathbf{u}))^2 dC_n(\mathbf{u})$$

where n is the number of observations in the model, d is the dimension of the model and θ_n is the parameter estimates of the proposed copula model.

Genest et al. conducted a comprehensive study of the various goodness of fit tests for copulas and concluded that the empirical based Cramer-von Mises (CvM) test is one of the best in comparison to others (i.e. test based on Rosenblatt transforms, Kendall transforms) in terms

Table 1: Summary of Cramer-von Mises Test

	Normal	t	Gumbel	Clayton
12 stocks (2012-17)	0.34	**0.584**	0	0.432
Financial	**0.984**	0.624	0.284	0.536
Tech	0.076	0.72	**0.968**	0.66
Industrial	0.52	**0.876**	0.068	0.18
Health	0.704	0.28	**0.824**	0.352
High Vol (2007-10)	0.12	**0.91**	0	0

of consistency and implementation. We implemented this test by using a double parametric bootstrap method as described in the aforementioned paper.

We tested the copulas using a significance level of 0.05; the higher the value, the better the fit. If the value was below 0.05, we rejected the null hypothesis, i.e. the model is not a good fit. Table 1 summarizes the results of this test on each of the fitted copulas. We found that t-distribution fits the best for the normal and high volatility period. This was expected because returns tend to have fat tails and have different correlations with each other in the long term. For the sector level copulas, we expected Gumbel and Clayton to fit better than the elliptical copulas. But we found that elliptical copulas fit better in some instances (e.g. financial and industrial) as well. One possibility is that Archimedean copulas tend to perform poorly for multivariate models (Hofert 2012). We can see this clearly for the high volatility case.

While the CvM test is consistent because it considers all misspecifications of the proposed model, we noticed a disadvantage in that the test does not provide guidance on the exact misspecifications. An alternative is to run additional GoF tests with each measuring a direction of misspecification such as particular moments and dependencies (Patton 2012). While these tests may not be consistent, coupling them with the CvM test will give a comprehensive understanding of the strengths and weaknesses of the proposed model.

Simulating VaR from Fitted GARCH + Copula Model

Value at Risk (VaR) measure is an important financial tool that measures maximum loss of a portfolio within a given probability (normally 95% or $\alpha = 0.05$). Formally,

Definition: Given X, for the confidence level $\alpha \in (0, 1)$ we define the α-Value-at-Risk (α-VaR) as

$$VaR_\alpha = min(x : F_X(x) \geq \alpha)$$

Table 2: VaR Violations with alpha = 0.05

	Normal	t	Gumbel	Clayton
12 stocks (2012-17)	0.042	0.037	0.047	0.032
Sector Level	0.026	0.032	0.032	0.026
High Vol (2007-10)	0.048	0.044	0.056	0.042

It is important that VaR is neither underestimated nor overestimated. Underestimation can lead to a massive financial loss if there is a shock to the market and overestimation can lead investors to execute sub-optimal investment strategies. Taking inspiration from Stulajter (2009), we tested our overall model (GARCH+copula) by simulating VaR with $\alpha = 0.05$ and comparing them with actual returns of a portfolio with equal weights on all stocks a one year period (2017 to 2018 for normal volatility period and 2010 to 2011 for high volatility period). We defined VaR violations to be the percentage of times actual returns fall below the simulated VaR. More formally,

Definition: For confidence level $\alpha \in (0, 1)$ we define

$$\mathcal{V}_\alpha = \frac{1}{m} \sum_{i=1}^{m} (r_i < VaR_i)$$

where m = number of observation days in testing data, r_i = actual portfolio return on day i, and VaR_i = simulated portfolio VaR on day i.

We judged that models that have $\mathcal{V}$ close to 0.05 were the best fit for the data. Table 2 summarizes the $\mathcal{V}$ for each model. We found that GARCH + Normal and GARCH + Gumbel performed the best in predicting VaR for the normal and high volatility periods. GARCH + Gumbel was not expected to perform well for these periods because the CvM tests failed; however as mentioned before, this can be investigated using additional GoF tests. Perhaps this will uncover minor misspecifications that resulted in a failing CvM test but a good overall model. Moreover, the GARCH + normal copula model was not expected to perform well in the high volatility period because returns were heavily skewed, and fat-tailed during this period. This might be due to the low number of observations in the testing data. For further investigation, we recommend using testing data with more than one year of observations.

We also found the GARCH + copula model did not perform very well for sector-fitted copulas. This was perhaps because there were more than two assets in each sector. Patton (2012) writes about vine copula which fits pairwise copulas to assets and builds to a high dimensional copula model. This is another line of investigation that can be done on this particular data set.

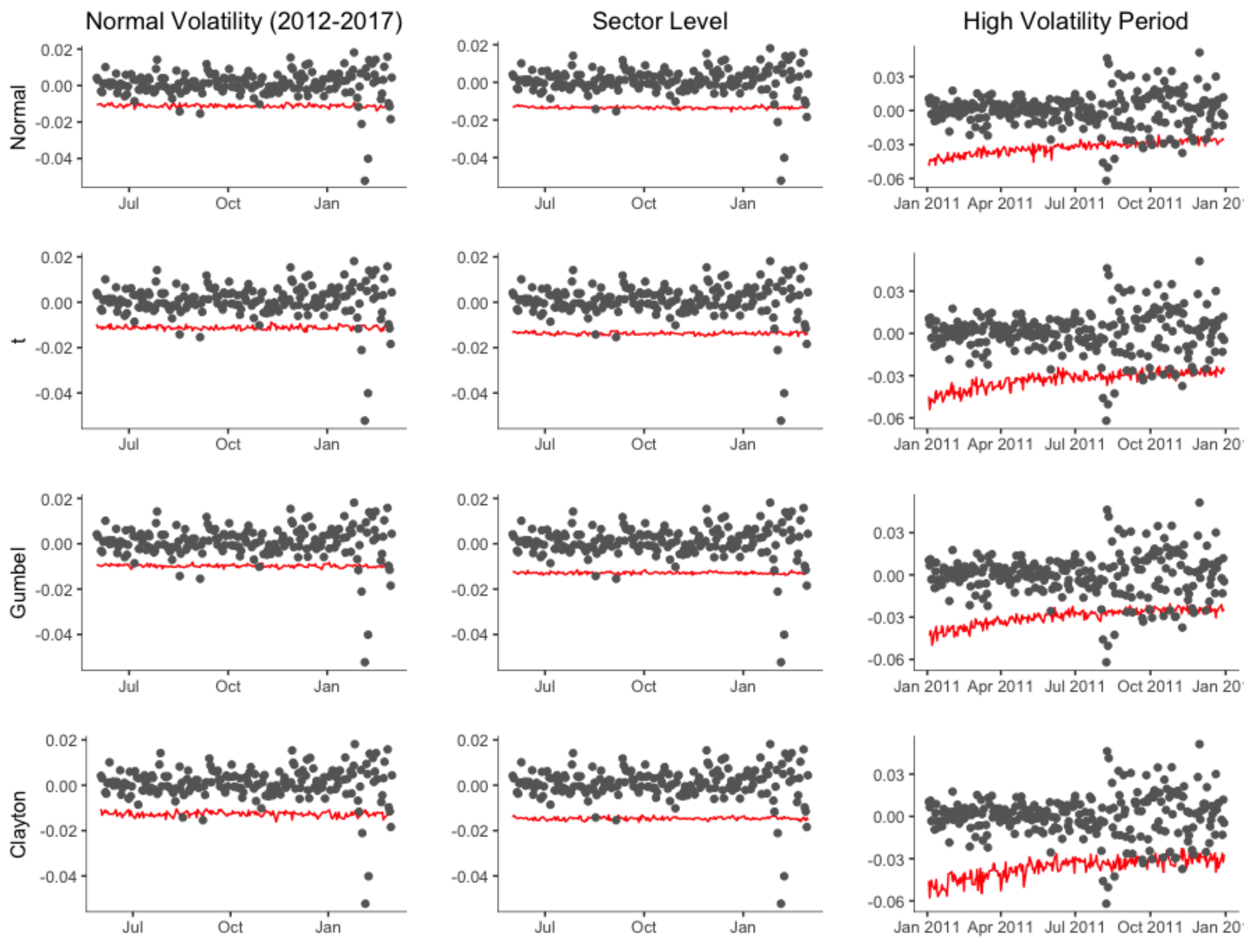

Figure 2: Estimated $VaR_{\alpha=0.05}$ and observed returns. The points indicate the observed returns, and the red lines indicates the estimated VaR level.

Conclusion

We implemented a GoF test for copula models that is consistent and easy to implement efficiently (computationally). However, we found the CvM test to be inconclusive by itself and needs to supplemented with further diagnostic tests as discussed. This will allow us to understand the strengths and limitations of each fitted copulas.

We considered VaR testing more significant because it considered the overall GARCH + copula and it involved a very useful financial metric. We found that GARCH with sector level copulas do not perform very well using this test. Our intention with this model was to be able to fit copulas easily to high dimensional models. Several publications show vine copula as a possible candidate for high dimensional modelling (see Patton 2012).

Normal and Gumbel performs the best in periods of normal volatility (2012-2017) and in periods of high volatility (2007-2010). However, we recommend using a larger number of observations in the testing data to make these findings conclusive.

References

Engle, R. (2001). GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics. Journal of Economic Perspectives, 15(4), 157–168. https://doi.org/10.1257/jep.15.4.157

Genest, C., R?millard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and Economics, 44(2), 199-213. https://doi.org/10.1016/j.insmatheco.2007.10.005

Hofert, M., Mächler, M., & McNeil, A. J. (2012). Likelihood inference for Archimedean copulas in high dimensions under known margins. Journal of Multivariate Analysis, 110, 133–150. https://doi.org/10.1016/j.jmva.2012.02.019

Nelsen, R. (2003). An Introduction to Copulas. Series in Statistics. https://doi.org/10.1007/b98966

Parra Palaro, H., & Koodi Hotta, L. (2006). Using Conditional Copula to Estimate Value at Risk. Journal of Data Science, 4, 93-115. https://doi.org/10.2139/ssrn.818884

Patton, A. J. (2012). A review of copula models for economic time series. Journal of Multivariate Analysis, 110, 4–18. https://doi.org/10.1016/j.jmva.2012.02.021

Posedel, P. (2005). Properties and Estimation of GARCH(1,1) Model. Metodološki Zvezki, 2(2), 243–257.

Stulajter, F. (2009). Comparison of different copula assumptions and their application in portfolio construction. Ekonomick? Revue - Central European Review of Economic Issues, 12(4), 191-204. https://doi.org/10.7327/cerei.2009.12.04