Contents

List of abbreviations

CDF	Cumulative density function
DCC	Dynamic conditional correlation
GARCH	Generalized autoregressive conditional heteroskedasticity
MC	Monte Carlo
MLEs	Maximum likelihood estimates
PDF	Probability density function
QQ	Quantile-Quantile

Glossary of used symbols

r_{t+1}	Daily rate of return at time t+1
S_t	Closing price of asset at time t
S_{t+1}	Closing price of asset at time t+1
R_{t+1}	Log return of asset at time t+1
μ	Mean of normal distribution
σ	Standard deviation of normal distribution
σ^2	Long-run Average variance
σ_t^2	Today's variance
σ_{t+1}^2	Tomorrow's variance
R_t^2	Squared returns at time t
$E\left[\sigma_{t+1}^2\right]$	Expected moving average variance
$E\left[\sigma_{t+k}^2\right]$	Expected moving average variance on k-days
l_t	Likelihood
L	Joint likelihood
$\ln L$	Logarithm of joint likelihood
γ_t	Correlation matrix
$\gamma^{1/2}$	Matrix squared root
$\rho_{ij,t}$	Correlation between asset i and j
D_{t+1}	Standard deviation matrix
$z_{i,t}$	Innovation term

$r_1(p)$	Empirical quantile for asset 1
$\Phi^{-1}_{(i-0.5)/T}$	Standard normal quantiles
$\tilde{t}(d)$	Standard t distribution
$\Gamma(z)$	Euler's gamma function
$N(0,1)$	Univariate standardized normal distribution
$D(0,1)$	Univariate standardized non-normal distribution
$f_i(z_i)$	Marginal distributions
u	Vector with elements
$F_i(z_i)$	Cumulative density functions
$C(\bullet)$	Copula CDF
$c(\bullet)$	Copula PDF
ρ^*	Copula correlation
I_n	n-dimensional identity matrix
$t_{(d.\rho^*)}$	Multivariate t distribution
t^{-1}	Inverse of student t distribution
$V_{PF,t}$	Portfolio value at time t
$N_{i,t}$	Units of asset i
$S_{i,t}$	Price of asset i
$\check{z}_{i,t}$	Pseudo random of $z_{i,t}$
$\check{R}_{i,t}$	Pseudo random of daily asset returns
Φ	Standard normal density

List of figures

1 Introduction

In finance, a stock return is a basic measurement index of the stock market. From the mathematical statistic viewpoint, stock returns are usually assumed to follow the normality distribution.

But according to the empirical examinations shows that the tails of stock returns are fatter than the tails of the normal distribution.[1] That means normal distribution can not accurately describe stock returns by definition.

Therefore, it is necessary to find new distribution assumptions to research this problem, as the student t distribution. Of course, it must be tested to verify whether it is better than the normal distribution.

Besides, stock returns are considered as portfolio or asset returns in practice. The multivariate distribution function is more appropriate to large assets of return.[2] And a method is needed to find, which can link the univariate model of stock returns to a joint distribution model.

To sum up, there are questions: How to form the proper joint distribution of asset returns? What are the criteria for a good approach to constructing joint distribution of asset returns?

I will find different ways to construct joint distribution and carried out empirical analysis of sample data in R to solve those problems. The GARCH volatility model, Maximum likelihood estimation method, QQ plot, DCC model, and threshold correlation plot will be discussed as base models. Lots of figures and calculations are used to render results of questions.

The goal of this paper is to answer the above questions. And according to different results of the goodness-of-fit, the best approach to constructing joint distribution of asset returns will be received.

Finally, the conclusion of the object: empirically compare different approaches to constructing a joint distribution of asset returns and its significance is the most important.

---

[1] Christoffersen(2012), S. 121

[2] Christoffersen(2012), S. 195

2 Research design

2.1 How to approach joint distribution of asset returns

2.1.1 Multivariate normal distribution of asset returns

2.1.1.1 GARCH model of dynamic variance

Daily stock return can be expressed by r_{t+1} using daily closed prices of stock S_{t+1} and S_t. That is $r_{t+1} = (S_{t+1} - S_t)/S_t$.

When the object of research is over a while and continuously, log returns of stock will be calculated. The notation R_{t+1} shows daily log return from natural logarithm: $R_{t+1} = \ln(S_{t+1}/S_t)$.[3] The stock log returns will be used as main observation objects of this paper.

The normal distribution called also Gaussian distribution in the point of mathematical probability theory. The formula is $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2)$ with notation $X \sim N(\mu, \sigma^2)$. And parameter μ is mean of normal distribution and σ means standard deviation. Notation σ^2 is the variance of distribution.

Standard normal distribution is the simplest form of normality distribution. Parameter μ is equal to 0 and σ is equal to 1. In this case, the formula defined as $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)$. When the sample Observations of a random variable with mean equal to zero and variance equal to one, that means the sample is converged to standard normal distribution.[4]

Firstly, the assumption of normal distribution for stock returns can be written as: $R_{t+1} = \sigma_{t+1} z_{t+1}$, where z_{t+1} is a innovation term and conform to normal distribution with $z_{t+1} \sim N(0,1)$.

[3] Christoffersen(2012a), S. 8; Christoffersen(2012b), S. 68

[4] Patel(1996), S. 19; Read(1996), S. 19

The above formula of assumption contains the relationship between dynamic variance σ_{t+1}^2 and the whole returns distribution. Given the time-varying variance: $\sigma_{t+1}^2 = \frac{1}{m}\sum_{\tau=1}^{m} R_{t+1-\tau}^2$, where m is random samples of objects. The notation R^2 is squared returns. It is an equivalent of R_{t+1} squared. A simple variance forecasting is measured by squared return.[5]

So that it is necessary to illustrate a generic variance model called JP Morgan's RiskMetrics. Tomorrow's variance can be calculated by all of the past squared stock returns, that is:

$$\sigma_{t+1}^2 = (1-\lambda)\sum_{\tau=1}^{\infty} \lambda^{\tau-1} R_{t+1-\tau}^2$$

And the actual variance with past squared returns can be explained by form：

$$\sigma_t^2 = (1-\lambda)\frac{1}{\lambda}\sum_{\tau=2}^{\infty} \lambda^{\tau-1} R_{t+1-\tau}^2$$

If combine these two formulas, tomorrow's variance will be written as:

$$\sigma_{t+1}^2 = \lambda\sigma_t^2 + (1-\lambda)R_t^2$$

It is clear to see that change of forecasting tomorrow's variance is generally same as the change of current variance and squared returns. Parameter λ is the only unknown value.[6] The process of estimation will be simplified. But this model ignores the leverage effect, which means a negative correlation between variance and stock returns.[7] And the RiskMetrics does not allow long horizons. If we considering brief dynamic variance, the result of the RiskMetrics and GARCH volatility model are similar to each other. But if time series is longer, today's variance will affect forecasting variance greatly. For the GARCH model, the predicted tomorrow's variance tends to the average value.[8] So GARCH variance model is a better tool to solve the problem.

[5] Christoffersen(2012), S. 68
[6] Christoffersen(2012), S. 69-70
[7] Christoffersen(2012), S. 76
[8] Christoffersen(2012), S. 72

For a series of logarithmic stock returns R_t^2 , the GARCH model for tomorrow's variance can be represented as $\sigma^2{}_{t+1} = \omega + \alpha R_t^2 + \beta\sigma_t^2$ (Formula 1), when the conditional $\alpha + \beta < 1$ is satisfied. And the GARCH model is flexible by setting parameter ω,α, and β.

Then the moving unconditional average variance can be defined in terms of the expected values by:

$\sigma^2 \equiv E[\sigma^2{}_{t+1}] = \omega + \alpha E[R^2{}_t] + \beta E[\sigma^2{}_t] = \omega + \alpha\sigma^2 + \beta\sigma^2$ (Formula 2).

Thus, it is easy to get a form from formula 1 and 2:

$\sigma^2{}_{t+1} = \sigma^2 + \alpha(R^2{}_t - \sigma^2) + \beta(\sigma^2{}_t - \sigma^2)$.

The GARCH model forecasting variance of log returns on k days can be written as:

$E_t[\sigma^2{}_{t+k}] = \sigma^2 + \alpha E_t[R^2{}_{t+k-1} - \sigma^2] + \beta E_t[\sigma^2{}_{t+k-1} - \sigma^2] = (\alpha + \beta)(E_t[\sigma^2{}_{t+k-1}] - \sigma^2)$.

To make it easier to modelling long-run forecasting, the form can be expressed by:

$E_t[\sigma^2{}_{t+k}] - \sigma^2 = (\alpha + \beta)^{k-1}(E_t[\sigma^2{}_{t+1}] - \sigma^2) = (\alpha + \beta)^{k-1}(\sigma^2{}_{t+1} - \sigma^2)$.[9]

From the above formula, we can see the advantage of the GARCH variance model is that the forecast of k days or months variance will be directly by tomorrow's variance $\sigma^2{}_{t+1}$ established. When the parameter $\alpha + \beta$ close to 1, the formula is simpler than before. That is $E_t[\sigma^2{}_{t+k}] = \sigma^2{}_{t+1}$.

On the other hand, the forecast GARCH variance model of K-days cumulative returns is:

$$\sigma^2_{t+1;t+k} = \sum_{k=1}^{K} \sigma^2_{t+1} = K\sigma^2 + \sum_{k=1}^{K} (\alpha + \beta)^{k-1}(\sigma^2_{t+1} - \sigma^2)$$

The K-days cumulative stock returns can be given by: $R_{t+1;t+K} = \sum_{k=1}^{K} R_{t+k}$. And

[9] Christoffersen(2012), S. 70-71

Variance in cumulative returns is $\sigma^2_{t+1:t+K} = \sum_{k=1}^{K} E_t\left[\sigma^2_{t+k}\right]$.[10]

2.1.1.2 Maximum likelihood estimation

According to the time-varying variance dependence on parameter ω,α, and β. So that how to estimate unknown parameters in the GARCH volatility model must be considering. So the maximum likelihood estimation method can be used to get those parameters.[11]

The maximum likelihood(MLEs) is a tool for inference and prediction of the target model.

The maximum likelihood estimated using the log of likelihood function because Log-likelihood is more close to its maximum value. The maximum values are approximate value and quantified by ratios of likelihood or log function scalars.[12]

Now recall the assumption of stock returns:

$R_t = \sigma_t z_t$, with $z_t \sim N(0,1)$

The likelihood function of assumption return can be written by:

$l_t = 1/\sqrt{2\pi\sigma_t^2}\exp(-\frac{R_t^2}{2\sigma_t^2})$, where notation l_t is a likelihood.

For multivariate likelihood of returns, the form can be written by:

$$L = \prod_{t=1}^{T} 1/\sqrt{2\pi\sigma_t^2}\exp(-\frac{R_t^2}{2\sigma_t^2})$$

The result of the maximum logarithm of the joint likelihood function will be same as the maximum value of likelihood estimation so that the formula is:

$$MaxInL = Max\sum_{t=1}^{T}\ln(l_t) = Max\sum_{t=1}^{T}\left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_t^2) - \frac{1}{2}\frac{R_t^2}{\sigma_t^2}\right]$$

This formula shows the way to evaluate unknown parameters ω,α, and β of the

---

[10] Christoffersen(2012), S. 72

[11] Christoffersen(2012),S. 73

[12] Maindonald, J.(2010), S. 133; Braun, W.(2010), S. 133

GARCH variance model, which can be represented as $\sigma^2{}_{t+1} = \omega + \alpha R_t^2 + \beta\sigma_t^2$. That means the term $-\frac{1}{2}\ln(2\pi)$ has seemed as constant parameter ω. Therefore, the simplest form is:

$$Max\sum^{T}\left[-\frac{1}{2}\ln(\sigma_t^2) - \frac{1}{2}\frac{R_t^2}{\sigma_t^2}\right]$$

There is impossible if choosing infinity large past returns as sample data. But let samples in sufficient numbers and monthly frequently is necessary for the calculation of parameters of the GARCH volatility model.[13]

For multivariate asset returns, the maximum log-likelihood function of the multivariate normal distribution can be written as:

$\ln(L) = -\frac{1}{2}\sum_t(\log|\gamma_t| + z_t'\gamma_t^{-1}z_t)$, with asset returns z_t and the determinant of the correlation matrix γ_t.

The function can be expressed using the idea of DCC model and the computational procedure will be easier. The maximum likelihood function can be written as:

$\ln(CL) = -\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j>i}(\ln(1-\rho_{ij,t}^2) + \frac{z_i^2 + z_j^2 - 2\rho_{ij}z_iz_j}{1-\rho_{ij}^2})$, where notation $\rho_{ij,t}$ means the correlation between different assets.[14]

The above calculation procedure is easy to run in the R program language. For example, unknown parameter omega, alpha, and beta in the GARCH variance model will be obtained as coefficients in console.

2.1.1.3 Dynamic conditional correlation(DCC) model of asset returns

The multivariate normal distribution can be obtained by univariate normal distribution with n-dimensions. In normal distribution, the notation of probability function is

[13] Christoffersen(2012),S. 74
[14] Christoffersen(2012),S. 164-165

$X \sim N(\mu, \sigma^2)$.

Now constructing a multivariate normal distribution of random vector $X = (x_1, x_2......x_n)^T$. The notation of multivariate probability function is $X \sim N(\mu, \sum)$ with mean of multivariate normal distribution μ and matrix of covariance ∑. The formula can be written as:

$$f_X(x_1,....,x_n) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp(-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu))$$

To discuss multivariate model for stock returns, the correlation between different stocks will be considering. So the dynamic conditional correlation(DCC) model is proposed. Also, the GARCH model is usually combined with the DCC model to analyze time-varying correlations.

In the bivariate case, the correlation between stock i and j is defined by notation $\rho_{ij,t+1}$. And the form of correlation is $\rho_{ij,t+1} = \frac{\sigma_{ij,t+1}}{(\sigma_{i,t+1}\sigma_{j,+1})}$, with variance and covariance of assets.

Then building the value of standard deviation matrix and correlation matrix using D_{t+1} and γ_{t+1}:

$$\sum_{t+1} = D_{t+1}\gamma_{t+1}D_{t+1} = \begin{bmatrix} \sigma_{1,t+1} & 0 \\ 0 & \sigma_{2,t+1} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12,t+1} \\ \rho_{12,t+1} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{1,t+1} & 0 \\ 0 & \sigma_{2,t+1} \end{bmatrix}$$

Now define $z_{i,t+1}$ as the standardized return of asset and connect it with each stock $R_{i,t+1}$ and dynamic variance $\sigma_{i,t+1}$: $z_{i,t+1} = \frac{R_{i,t+1}}{\sigma_{i,t+1}}$.

The formula for conditional correlation using covarince of standardized returns to come true.[15] That is $E_t(z_{i,t+1}, z_{j,t+1}) = E_t((\frac{R_{i,t+1}}{\sigma_{i,t+1}})(\frac{R_{j,t+1}}{\sigma_{j,t+1}})) = \frac{\sigma_{ij,t+1}}{\sigma_{i,t+1}\sigma_{j,t+1}} = \rho_{ij,t+1}$

For bivariate correlated stocks, the stock with the correct correlation matrix is:

[15] Christoffersen(2012),S. 159-160

$$E\left[z_{t+1}(z_{t+1})'\right] = \begin{bmatrix} 1 & \rho_{1,2} \\ \rho_{1,2} & 1 \end{bmatrix} = \gamma$$

Where the bivariate correlation matrix can also be written as:

$$\gamma = \gamma^{1/2}(\gamma^{1/2})'$$

So that the matrix squared root $\gamma^{1/2}$ can be constructed by form:

$$\gamma^{1/2} = \begin{bmatrix} 1 & 0 \\ \rho_{1,2} & \sqrt{1-\rho_{1,2}^2} \end{bmatrix}$$

The standard normal variable in the bivariate case will be defined as:

$z_1 = z_1^u$ and $z_2 = \rho_{1,2} z_1^u + \sqrt{1-\rho_{1,2}^2}\, z_2^u$

Define notation z_1 as vector of returns for asset 1 and z_2 for asset 2.[16] And these formulas will be used later when building threshold correlation plots.

For asset z_1 and z_2, the probability density function in this two dimensions case is

$$f(z_1, z_2) = \frac{1}{2\pi\sigma_{z1}\sigma_{z2}\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(z_1-\mu z_1)^2}{\sigma_{z1}^2} + \frac{(z_2-\mu z_2)^2}{\sigma_{z2}^2} - \frac{2\rho(z_1-\mu z_2)(z_2-\mu z_2)}{\sigma_{z1}\sigma_{z2}}\right]\right)$$

The parameter ρ is the correlation between two assets.[17]

To make the density formula easier define and modeling in the R program, the function can be given by the following form:

$f(z_1, z_2) = \Phi_\rho(z_1, z_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\frac{z_{1,t}^2 + z_{2,t}^2 - 2\rho z_1 z_2}{2(1-\rho^2)}\right)$, where the bivariate correlation matrix between two assets is $|\gamma| = \begin{vmatrix} 1 & \rho \\ \rho & 1 \end{vmatrix} = 1-\rho^2$.[18] The correlation ρ has been already computed by DCC model.

The n-dimensions multivariate probability density function is:

[16] Christoffersen(2012),S. 184

[17] Tong(1990), S. 6-7

[18] Christoffersen(2012), S. 196

$$f(z_t;\gamma)=\Phi_\gamma(z_t)=\frac{1}{(2\pi)^{n/2}|\gamma|^{1/2}}\exp(-\frac{1}{2}z_t^{'}\gamma^{-1}z_t)$$.[19]

2.1.1.4 Threshold correlation plot for multivariate normal distribution

Now modeling multivariate normal distribution for asset returns. The graph method threshold correlation is intuitive to show the relationship between asset returns and normal distribution in the multivariate case. It describes not only the relationship between different shocks and also the tail shape of multivariate distribution.

Firstly, a correlation between stocks in the bivariate case with probability must be mentioned. Considering probability 0.5 as a cut-off point and $z_{1,t}$, $z_{2,t}$ as vector of asset standard returns for two companies. And there is a function of asset returns: $z_t = r_t / \sigma_t$.

When the probability smaller than 0.5, the two assets are both lower than there percentile values. When the probability bigger than 0.5, the two assets are both higher than there percentile values. Notation $r_1(p)$ and $r_2(p)$ are empirical quantile for asset 1 and asset 2.

The following form of threshold correlation for assets with probability p can be written as:

$$\rho(r_{1,t}, r_{2,t}, p) = \begin{Bmatrix} Corr(r_{1,t} \le r_1(p) and r_{2,t} \le r_2(p)), p \le 0.5 \\ Corr(r_{1,t} > r_1(p) and r_{2,t} > r_2(p)), p > 0.5 \end{Bmatrix}$$

If the two lines are matched well, that means the multivariate normal distribution conforms with asset returns.

The threshold correlation describes also dependence of asset returns. The observation stocks could be both negative or both positive.[20] Then analyze them with probability. And combine multivariate normal density function with threshold correlation plot to

[19] Christoffersen(2012), S. 197

[20] Christoffersen(2012),S. 194-195

show bivariate normal distribution of asset returns for different values of correlation. If lines show flexible degrees of tail dependence between two variables in the threshold correlation from the bivariate normal distribution, or rather, lines can accurately describe features of data, that means the joint distribution model is fit to asset returns.

2.1.2 The multivariate student t distribution of asset returns

2.1.2.1 QQ plot of Non-normality and standard t distribution

The Quantile-Quantile(QQ) plot compares empirical quantiles of distributions. It gives a way to find whether a distribution model fits to the distribution of random variables.[21]

At first, QQ plot will be used to show the Non-normality of stock returns. The diagonal line $y = x$ is the quantiles of normal distribution, if the quantiles function of stock returns is normally distributed, it will be tended to this diagonal line.[22]

From the form of stock returns before, we have known that $R_{PF,t} = \sigma_{PF,t} z_t$ for rate of portfolio return. Building z_t as standard normal distributed and sort these values. That is $z_t = \dfrac{R_{PF,t}}{\sigma_{PF,t}}$.

Then find the probability below than $\dfrac{i-0.5}{T}$. The value -0.5 is an adjustment for those random variables. And T is all of the sample objects.

Define standard normal quantiles as function $\Phi^{-1}_{(i-0.5)/T}$. The meaning of Axis is same as the definition as before. The Axis of quantiles of the normal distribution can be given by:

---

[21] Beirlant, J.(2004), S3-4; Geogebeur, Y.(2004), S.3-4; Teugles, J.(2004), S. 3-4; Segers J.(2004), S. 3-4;Waal, D.D.(2004), S. 3-4; Ferro, C.(2004), S. 3-4

[22] Christoffersen(2012),S. 123

$\{X_i, Y_i\} = \{\Phi^{-1}_{(i-0.5)/T}, z_i\}$.[23]

Secondly, combine normally distributed GARCH model and t distributed GARCH model with QQ plot together to find that QQ plot has smaller deviations from student t distribution. So that trying to approach multivariate t distribution of asset returns.

2.1.2.2 Threshold correlation plot for multivariate student t distribution

The Student t distribution called also t distribution for short, it used to calculate the mean of all populations with normal distribution and an unknown value of deviation.

$\tilde{t}(d)$ represent the notation of standard t distribution. The key advantage of t distribution is that describing features of stock returns better than normal distribution.

The function of t distribution is $f(t) = \dfrac{\Gamma(\frac{d+1}{2})}{\Gamma\frac{d}{2}\sqrt{d\pi}}(1+\dfrac{t^2}{d})^{-\frac{(1+d)}{2}}$, with $d > 0$. The parameter d is the number of degrees of freedom and $\Gamma(z)$ is Euler's gamma function.

Then considering standardized t distribution $\tilde{t}(d)$ of the above function. That is

$$f(t) = \frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})\sqrt{\pi(d-2)}}(1+\frac{(t/\sqrt{d/(d-2)})^2}{(d-2)})^{-\frac{(1+d)}{2}}, \text{ with } \frac{t}{\sqrt{d/(d-2)}} = z \text{ for } d > 2.$$

The random term z obtains the value of mean equal to zero and standard deviation equal to one.[24]

Using the maximum log-likelihood function to estimate the unknown parameter d：

$$\ln L = \sum_{t=1}^{T} \ln(f_{\tilde{t}(d)}(z_t;d)) = T\{\ln(\Gamma((d+1)/2)) - \ln(\Gamma(d/2) - \ln(\pi)/2 - \ln(d-2)/2\}$$
$$-\frac{1}{2}\sum_{t=1}^{T}(1+d)\ln(1+(R_{PF,t}/\sigma_{PF,t})^2/(d-2)).$$

[23] Christoffersen(2012),S. 124

[24] Christoffersen(2012), S. 128

The variance $\sigma^2_{PF,t}$ has been modeled by GARCH model before, therefore, the parameter d effects maximum likelihood $\ln L$.[25]

Notice that multivariate t distribution with the correlation ρ between those asset returns, the bivariate t distribution can be written as:

$$f(z_1, z_2) = \frac{\Gamma((d+2)/2)}{\Gamma(d/2)(d-2)\pi(1-\rho^2)^{1/2}}(1+\frac{z_1^2+z_2^2-2\rho z_1 z_2}{(d-2)(1-\rho^2)})^{\frac{-(d+2)}{2}}$$

In the bivariate case, the probability density function of multivariate t distribution is

$$f(z_1, z_2) = \frac{\left|\Sigma^{-1}\right|^{\frac{1}{2}}}{2\pi}(1+\sum_{i,j=1}^{2,2}\frac{\Sigma^{-1} z_1 z_2}{d})^{\frac{-(d+2)}{2}}$$, for all $d > 2$.

Finally, the n-dimensions multivariate t distribution will be defined as:

$$f(z; d, \gamma) = \frac{\Gamma((d+n)/2)}{\Gamma(d/2)((d-2)\pi)^{n/2}\left|\gamma\right|^{1/2}}(1+\frac{z'\gamma^{-1}z}{d-2})^{-\frac{(d+n)}{2}}$$, with $d > 2$.[26]

And constructing multivariate t density function with threshold correlation plot to show bivariate t distribution of asset returns for different values of correlation. If lines can accurately describe features of data, that means the joint distribution model is fit to asset returns.

2.1.3 Copulas model for joint distribution of asset returns

2.1.3.1 Theoretical foundation: Sklar's Theorem

Building the univariate standardized non-normal distribution using $R_{PF,\ t} = \sigma_{PF,t} z_t$,

that is $z_t = \frac{R_{PF,t}}{\sigma_{PF,t}}$, which contains forecast conditional variance $\sigma_{PF,t}$ and univariate distribution D(0,1) with mean of zero and deviation of one.[27]

The notation can be written as $z_t \sim i.i.d.D(0,1)$.

[25] Christoffersen(2012), S. 129-130
[26] Christoffersen(2012), S. 198-199
[27] Christoffersen(2012),S. 123

Therefore, the multivariate non-normal distribution for return of asset or portfolio will be defined as $z_t \sim D(0,\gamma_t)$. The connotation of notation z_t in this situation means vector of varied stocks asset and γ_t as the matrix of time varying correlation.

Multivariate distribution functions are not enough for extreme value and have shortcomings if the precondition is non-normality.[28]

So that copula models offer a way to combine individual variables or univariate distribution into a joint model. The process of modeling copulas in this research is to make the marginal probability functions form a joint distribution.[29] If the notation $f_i(z_i)$ is returns of asset i(i=1,2......n), u_i is the probability cumulative density function(CDF) with form $u_i = F_i(z_i)$ for marginal distributions of n assets.[30]

Sklar's Theorem is the basis of copulas. It offers a connection between marginal variables and the joint distribution and constructed by $C(x_1,...,x_n)$. Different individual models can be together into a joint distribution.

In the bivariate case, there is standard joint distribution function $F_{x,y}$ with two CDFs:

$F_{x,y}(x,y) = C(F_x(x), F_y(y))$.

For multivariate standard returns z_i, the copula can be written by:

$F(z_1,...,z_n) = C(F_1(z_1),...,F_n(z_n)) = C(u_1,...,u_n)$, with $u_1 = F_1(z_1)$

Sklar's Theorem for multivariate probability density function will be given by form:

$$f(z_1,...,z_n) = \frac{\partial^n C(F_1(z_1),...,F_n(z_n))}{\partial z_1 ... \partial z_n} = \frac{\partial^n C(u_1,...,u_n)}{\partial u_1 ... \partial u_n} \times \prod_{i=1}^{n} f_i(z_i) = c(u_1,...,u_n) \times \prod_{i=1}^{n} f_i(z_i)$$

Where n-dimensions marginal distribution for asset returns $f_1(z_1),...,f_n(z_n)$ and the notation $c(u_1,...,u_n)$ as probability density copula defined.

The log-likelihood function of copula PDF noted as:

[28] Christoffersen(2012),S. 193

[29] Kole, Erik(2006): Journal of Banking & Finance, in: 31(2007), S. 2407

[30] Christoffersen(2012), S. 203

$$\ln L = \sum_{t=1}^{T} \ln c(u_1^t, ..., u_n^t)$$

It gives a possibility to modeling joint distribution of different assets, which can use different univariate distribution and estimating its parameters.[31]

2.1.3.2 The normal copula of asset returns

Normal copula called also Gaussian copula. The joint distribution from dependence multivariate standard normal function using normal copula, where γ^* means a matrix of correlation between different inverse cumulative density functions of individual normal distribution.

In the bivariate case, the normal copula for two normal inverse cumulative density functions is:

$C(u_1, u_2; \rho^*) = \Phi_{\rho^*}(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) = \Phi_{\rho^*}(\Phi^{-1}(F_1(z_1), \Phi^{-1}(F_1(z_2)))$, with probability CDF of returns $u_i = F_i(z_i)$ and probability correlation of two asset ρ^*.

Notice that $F_1(z_1)$ and $F_2(z_2)$ are marginal distributions for standard returns of asset 1 and asset 2.[32]

And the normal copula of probability density function can be given by the following form:

$$c(u_1, u_2; \rho^*) = \frac{1}{\sqrt{1-\rho^{*2}}} \exp(-\frac{\Phi^{-1}(u_1)^2 + \Phi^{-1}(u_2)^2 - 2\rho^* \Phi^{-1}(u_1)\Phi^{-1}(u_2)}{2(1-\rho^{*2})}$$
$$+\frac{\Phi^{-1}(u_1)^2 + \Phi^{-1}(u_2)2}{2})$$

The function of maximum log-likelihood is:

$$\ln L = \sum_{t=1}^{T} \ln c(u_1, u_2) = -\frac{T}{2}\ln(1-\rho^{*2}) - \sum_{t=1}^{T} \frac{\Phi^{-1}(u_1)^2 + \Phi^{-1}(u_2)^2 - 2\rho^* \Phi^{-1}(u_1)\Phi^{-1}(u_2)}{2(1-\rho^{*2})}$$
$$+\frac{\Phi^{-1}(u_1)^2 + \Phi^{-1}(u_2)^2}{2}$$

[31] Christoffersen(2012), S. 203-204
[32] Christoffersen(2012), S. 205

In n-dimensions normal copula, the multivariate cumulative density functions with n asset can be written as:

$$C(u_1,...,u_n;\gamma^*) = \Phi_{\gamma^*}(\Phi^{-1}(u_1),...,\Phi^{-1}(u_n))$$

The multivariate probability density function is:

$c(u_1,...,u_n;\gamma^*) = \left|\gamma^*\right|^{-\frac{1}{2}} \exp(-\frac{1}{2}\Phi^{-1}(u)'(\gamma^{*-1} - I_n)\Phi^{-1}(u))$, with an identity matrix

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Log-likelihood can be used for estimate correlation matrix γ^*, the function is:

$$\ln L = \sum_{t=1}^{T} \ln c(u_{1,t},...,u_{n,t}) = -\frac{1}{2}\sum_{t=1}^{T} \ln\left|\gamma^*\right| - \frac{1}{2}\sum_{t=1}^{T} \Phi^{-1}(u_t)'(\gamma^{*-1} - I_n)\Phi^{-1}(u_t)$$

The correlation matrix of normal copula can be written as:

$\gamma^* = \frac{1}{T}\sum_{t=1}^{T} z_t^* z_t^{*'}$, where notes $z_{i,t}^*$ as copula return for asset i on day t and defined by the formula: $z_{i,t}^* = \Phi^{-1}(F_i(z_{i,t}))$.

The key difference between multivariate normal distribution and normal copula is that the normal copula can be written by non-normal marginal distributions not only normal marginal distributions. So normal copula is more convenient to calculation and analysis than normal distribution.

But there is the same disadvantage as normal distribution. The normal copula can not building multivariate model for asset returns with extreme variables, which have high correlations of tails.[33]

To show the shortcoming of normal copula and compare it with the t copula model. Considering the Kendall Tau method to estimate based parameter and build contour plots of normal copula and t copula.

[33] Christoffersen(2012), S. 206-207

2.1.3.3 The t copula of asset returns

According to contour plots of these copula models, assume that the t copula model is better for modeling joint distribution of asset returns.

Now considering the student t copula. The t copula model is more flexible for modeling multivariate distribution with extreme variables than the normal copula.

For two asset returns cumulative density functions, the t copula can be given by the following form with parameter d for degrees of freedom:

$$C(u_1,u_2;\rho^*,d)=t_{(\rho^*,d)}(t^{-1}(u_1;d),t^{-1}(u_2;d))$$

The notation $t_{(d,\rho^*)}$ states multivariate t distribution with parameter d and correlation ρ^*. And t^{-1} is the inverse of student t distribution function.

For multivariate asset returns, the n-dimensions cumulative density function of t copula can be written as:

$C(u_1,...,u_n;\gamma^*,d)=t(d,\gamma^*)(t^{-1}(u_1;d),...,t^{-1}(u_n;d))$, with a matrix of correlation γ^*.

The probability density function in the bivariate case is:

$$c(u_1,u_2;\rho^*,d)=\frac{\Gamma((d+2)/2)}{\sqrt{1-\rho^{*2}}\Gamma(d/2)}(\Gamma(d/2)/\Gamma((d+1)/2))^2$$

$$\times\frac{(1+\frac{(t^{-1}(u_1,d))^2+(t^{-1}(u_2,d))^2-2\rho^*t^{-1}(u_1)t^{-1}(u_2)}{d(1-\rho^*)^2})^{-(d+2)/2}}{(1+\frac{(t^{-1}(u_1;d))^2}{d})^{-(d+1)/2}(1+\frac{(t^{-1}(u_2;d))^2}{d})^{-(d+1)/2}}$$

And the multivariate PDF of t copula obtain the formula:

$$c(u_1,...,u_n;\gamma^*,d)=\frac{\Gamma((d+n)/2)}{\left|\gamma^*\right|^{1/2}\Gamma(d/2)}(\frac{\Gamma(d/2)}{\Gamma((d+1)/2)})^n\frac{(1+\frac{1}{d}t^{-1}(u;d)'\gamma^{*-1}t^{-1}(u;d))^{-(d+n)/2}}{\prod_{i=1}^{n}(1+\frac{(t^{-1}(u;d))^2}{d})^{-(d+1)/2}}$$

For estimating parameter d and correlation matrix of t copula, the maximum log-likelihood form is:

$$\ln L=\sum_{t=1}^{T}\ln c(u_{1,t},...,u_{n,t}).$$

Here the copula shocks of t copula will be defined as:

$$z_{i,t}^{*} = t^{-1}(F_i(z_{i,t}); d)$$

And the form of the correlation matrix is similar to the formula of correlation matrix for normal copula.[34]

Finally, plot the scatter plot of sample observations and simulated data under normal marginals distribution functions and student t copula to verify whether the t copula model fits to multivariate asset returns. If most of the simulated observations are fit to sample data i.e. these points coincide with each other, which means t copula a good approach to constructing joint distribution of asset returns.

2.2 Evaluation criteria of approaches

2.2.1 The univariate models of individual asset

The univariate model of each asset is basis before considering multivariate model of asset returns. There are two steps of univariate models.

Forecasting tomorrow's variance of asset returns is the first part of this model. Define daily asset log returns and assume daily returns normal distributed. Then estimating dynamic variance by squared return. To catch features of each asset better, the GARCH volatility model will be used. This part shows the effect of daily asset returns to dynamic volatility.[35]

The second part of univariate models is modeling the non-normality of portfolio returns.

Recall the formula of portfolio: $V_{PF,t} = \sum_{i=1}^{n} N_{i,t} S_{i,t}$. Given log returns of the portfolio as: $R_{PF,t} = \ln(V_{PF,t} / V_{PF,t-1})$. And the assumption of conditional volatility under non-normal distribution can be written as $R_{PF,t} = \sigma_{PF,t} z_t$, with $z_t \sim D(0,1)$.[36]

---

[34] Christoffersen(2012), S. 208-209

[35] Christoffersen(2012), S. 67-68

[36] Christoffersen(2012), S. 123; Christoffersen(2012), S. 143

Constructing theoretical quantiles against sample quantiles QQ plot and compare the univariate normal distribution and the t distribution of asset returns using the GARCH model together.

At last, find the better approach of those distributions to fit portfolio data.

2.2.2 Monte Carlo simulation

The Monte Carlo simulation method describes distributions by simulated random samples. This process can be implied by drawing lots of artificial random samples and observing features of the distribution model.[37]

If we generate pseudo random of innovation term $z_{i,t}$. That can be written as:

$\check{z}_{i,1}$, with i=1,2,3,...,MC

So that the formula of daily asset returns and period (like K days) returns will be implied by:

$$\check{R}_{i,t+1} = \sigma_{t+1}\check{z}_{i,t+1} \quad \text{and} \quad \check{R}_{i,t+1,t+k} = \sum_{k=1}^{K}\check{R}_{i,t+k}, \quad \text{where i=1,2,3,...,MC}$$

The multivariate normal distribution and t distribution of asset returns with different correlations can be constructed by Monte Carlo random numbers. The process for multivariate Monte Carlo is: At first, given multivariate random $\check{z}^{u}{}_{i,1}$. And then using matrix square $\gamma^{1/2}$ to correlate those random data with the DCC model. Finally, assessing the threshold correlation from multivariate distributions models.

Notice that Monte Carlo simulation allows only for standardized distributed returns. The random number of samples is flexible.[38] To increase the accuracy of results, I will choose as much simulated data as possible.

[37] Christopher Z.(1997), S. 2-3

[38] Christopher Z.(1997), S. 177

3 Data

3.1 Data collection

In this paper, the daily closing price of Alibaba Group and Amazon from January 01, 2015 to January 01, 2020(five years of data) are initial sample data. These two companies were both listed in American. To make sure samples with the same monetary units is a prerequisite for analysis. Those data resources are directly given by Yahoo Finance in R. For example, the first closing price for Alibaba Group on January 02, 2015 is 103.60 and the price for Amazon is 308.52 at the same time.

And then is data cleaning of daily closed price for these two sample companies because there are much missing values. e.g. there are no closed prices on January 03, 2015 and January 04, 2015. The data cleaning process is necessary for the calculation of asset returns later.

So daily asset returns are observation data, which used to analyze with distribution functions and copula models together. The value of stock returns can be estimated by closing prices.

3.2 Data processing

Firstly, to get the sample data from Alibaba Group in five years and estimate daily log returns using the packet “quantmod” in R. In order to use closing prices from January 01, 2015 to January 01, 2020 as empirical data, the packet “zoo” and it’s extended implementation packet “xts” must be used.

In R, the packet “zoo” offers time series and is the basis of stock analysis. The packet “xts” enriches the functions of time series data processing.

So that it is easy to obtain daily open price, high price, low price, closing price, volume, and adjusted value of Alibaba Group. The 4^{th} column of the data is the closing price and will be chosen to calculate daily log returns in these five years via a function of log returns in R.

Then drawing the histogram of daily log returns. And add a line to the histogram, which describes normal distribution with the same mean and standard deviation as log returns in figure 1.

For squared returns and moving average variance of asset, the daily log returns will be calculated using the R library function.

And for moving average variance, the following packages needed to load: dplyr, stats and base. The packet "dplyr" is mainly used to data cleaning, including data select, filter, arrange, mutate and summarize, etc. According to the form of tomorrow's variance: $\sigma_{t+1}^2 = \frac{1}{m}\sum_{\tau=1}^{m} R_{t+1-\tau}^2$ with 10 observations. Setting the R library function "filter" with data of log returns to show the result. The time series is from 2015 to 2020 and the frequency of the data will be built as 12 months. Finally, the plot of time-varying variance can be constructed.

Contrast with moving average variance is the GARCH volatility model. The GARCH variance model for log returns can be achieved by package "fGarch" in R. The log-likelihood of unknown parameters will be shown in console called coefficient(s). And using normal distribution and student t distribution to construct the GARCH(1,1) model.

Then constructing line chart of dynamic volatility from fitted GARCH objects. Time will be same as moving average variance: From 2015 to 2020 with frequency 12 months.

R functions "qqnorm" and "qqline" are used for the normality test of Alibaba's returns.

Now considering a combination of two methods: GARCH volatility model and QQ plot. It offers an intuitive way to compare goodness-of-fit of standard normal distribution and t distribution.

Combine Alibaba's GARCH shocks against normal distribution at first. And compare it with the Alibaba's GARCH shocks against student t distribution. Number 13 means the graph type will be chosen as QQ plot. Finally, get QQ plots of normally distributed GARCH shocks and t distributed GARCH shocks in figure 7.

Fitting the DCC-GARCH model to time series of stock returns. The chart of the dynamic conditional correlation between Alibaba Group and Amazon will be completed.

The first step is to get closing prices from Amazon and calculate its daily log returns from January 01, 2015 to January 01, 2020. It is same as the process before that obtains data and estimates returns from Alibaba Group. To make data of two companies into a data frame. The name of the first column is daily returns i.e. Alibaba's returns and the second column is daily returns 1 i.e. Amazon's returns.

Next, loading package "rmgarch", package "rugarch" and the default package "parallel". The packet "rmgarch" handles multivariate data in the GARCH model. Assume that the bivariate data of stock returns are normally distributed. Then specify "mean.model" as simplest armaOrder(0,0). The "variance.model" set up a simple GARCH(1,1) in "sGARCH" model. Insert those models together. And connect the DCC model with the GARCH model under multivariate normal distribution in R. Given conditional correlations between asset returns from two companies with random number of observation 10. The conditional correlation plot will be received.

Figure 9 shows the threshold correlation for Alibaba Group versus Amazon and threshold correlation from normal distribution with correlation matching sample data. For the threshold correlation plots, the package "ggplot2" and "tidyr" are necessary.

The data frame of bivariate log returns from Alibaba and Amazon is a foundation. Combine the returns of these two companies into a data frame. Recall function of assets, quantile of assets, and probability according to the corresponding form:

$$\rho(r_{1,t}, r_{2,t}, p) = \begin{cases} Corr(r_{1,t} \leq r_1(p) and r_{2,t} \leq r_2(p)), p \leq 0.5 \\ Corr(r_{1,t} > r_1(p) and r_{2,t} > r_2(p)), p > 0.5 \end{cases}$$

Constructing correlation function, which contains the above conditions. Then setting function on random samples of normal distribution. Define repeat value equal to zero. The function of random variable 1 will be written as "data$r1" and the variable 2 will be constructed same as in the DCC model:

$data\$r1 * \rho + data\$r2 * \sqrt{(1-\rho^2)}$

Quantile's scalar states from 0.15 to 0.90 with equally increments 0.01. The threshold correlation of normal distribution and the threshold correlation with matching log returns will be defined. Given 30000 random numbers of normal distribution using the Monte Carlo simulation method.

Notice that if make percentile, the correlation between two asset returns, and correlation with normal distribution into a data frame, there are five columns of data will be represented. But only data on columns one, three, and five are useful. After that, i will build plots of threshold correlations with "ggplot" in R. So that these two lines can appear on the same graph.

In order to analyze the threshold correlation from multivariate standard normal distribution. Repeating zero and random simulations of normal distribution into the data frame.

Given value of random correlation ρ as -0.3, 0, 0.3, 0.6 and 0.9. Defining normal distribution function from Alibaba's returns and Amazon's returns with different values of correlations. Combine all of vectors of percentile and correlations. Selecting data on columns one, two, four, six, eight, and ten, which including percentile and corresponding correlations. Then drawing threshold correlation from multivariate normal distribution with "ggplot" in R.

Up to now, constructing threshold correlation implied by multivariate t distribution with different values of correlation. The process is similar to the threshold correlation from bivariate normal distribution. Repeating zero and random simulations of t distribution into the data frame. Given random correlations as -0.25, 0, 0.25, 0.75 and 0.9. Define the multivariate t distribution function of asset returns same as bivariate normal distribution because d is a scalar and it has no effect on sample data. The Monte Carlo random numbers are same as before i.e. 30000. Assign parameter d as 15 here is flexible. The line chart of threshold correlation against percentile by multivariate t distribution will be built.

Then constructing contour plot for normal copula and student t copula. The package “VineCopula” and “copula” are for copula models and acts on parameter estimation, simulation, model selecting, visualization, etc. Creating parameter of bivariate copula model with Kendall Tau method. Here I will define the same tau as 0.7 for bivariate normal copula and t copula. So that the two contours are comparable. Then determining the family of the objects. Number 1 means Gaussian(normal) copula and number 2 is the student t copula. For t copula, a second parameter ν can compete as 4. At last, building contour plots for those two copula models in R.

Finally, connect the copula model to sample data.

The first step is modeling marginal distributions. For the sake of simplicity, assume that marginal distributions of asset returns conform to normal distribution. Define parameters of marginal distributions i.e. “mean(ret)” and “sd(ret)” for Alibaba’s returns and “mean(retamazon)” and “sd(retamazon)” for Amazon’s returns. Apply the copula model in R and setting the copula model as t copula in the multivariate case. Then generate simulated observations with random samples 4000. Here shows again Monte Carlo simulation. At last, plot the scatter plot of sample observations and simulated data under normal marginals distribution functions and student t copula model. Add legends to make the scatterplot more clearly.

4 Empirical results

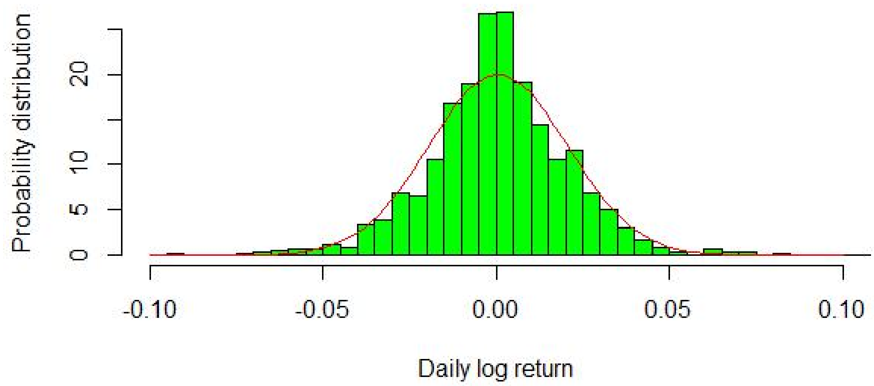
Before approaching the joint distribution model of stock returns. Recall the assumption for stock returns is: $R_{t+1} = \sigma_{t+1} z_{t+1}$, with $z_{t+1} \sim N(0,1)$

At first, using a histogram of daily Alibaba log returns and line of normal distribution to show the mismatch. The following green histograms are daily log returns for Alibaba Group from January 01, 2015 to January 01, 2020.

It is clear to indicate that histograms in green are higher than the red curve, which describes the normal distribution with the same mean and standard deviation as stock returns. And there is a fatter tail of log stock returns than the normal distribution. So the stock returns do not correspond to the density of normal distribution function.

This phenomenon gives us the motivation to continue research because figure 1 overturns the general assumption: stock returns are usually assumed to follow the normality distribution. In other words stock returns are non-normal distributed.[39]

Figure 1: Histogram of daily Alibaba log returns and normal distribution



Creating a forecasting tomorrow's variance model using squared return with the simple average of m objects, by laying out the form: $\sigma^2_{t+1} = \frac{1}{m}\sum_{\tau=1}^{m} R^2_{t+1-\tau} = \sum_{\tau=1}^{m}\frac{1}{m}R^2_{t+1-\tau}$.

The value of moving average variance dependence on tomorrow's squared returns. So the first step is to show time-varying squared returns. The following graph contains a line of daily squared returns for Alibaba Group from Jan 2015 to Jan 2020 in green. The obvious kurtosis is around the end of 2016 to 2017. From January 2015 to the end of 2019, the tend of most squared returns waves around 0.005.

[39] Christopher Z.(1997), S. 121

Figure 2: Squared returns for Alibaba from January 2015 to January 2020

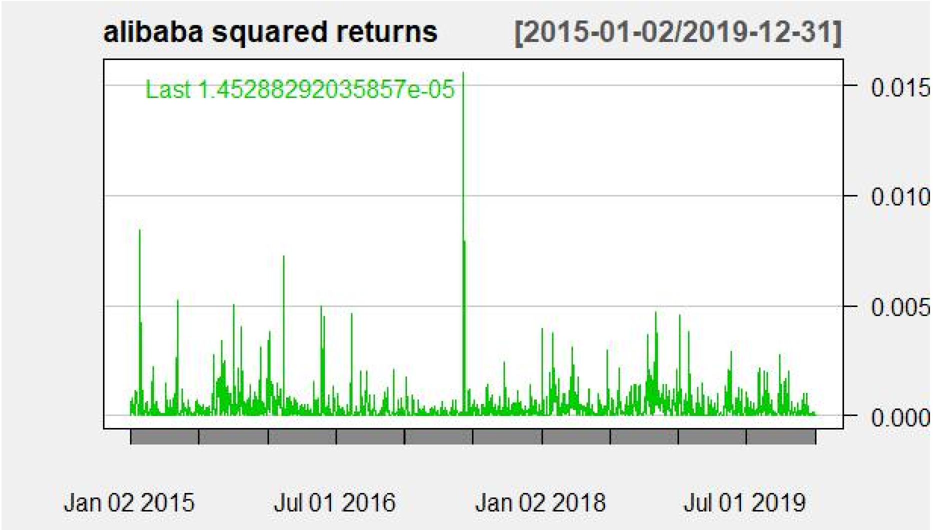


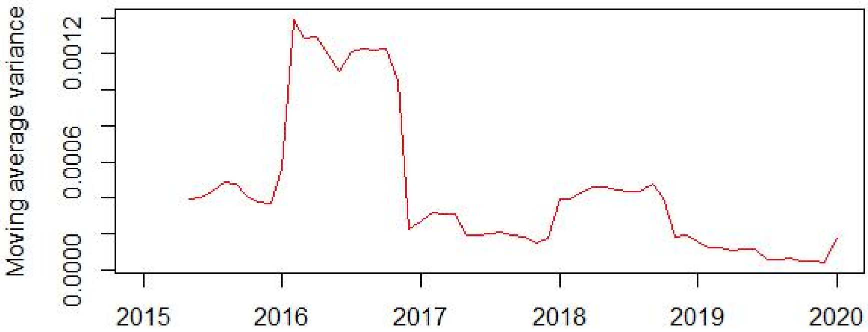
Figure 3 shows a line of moving average variance with ten random objects. The highest line segment is also around 2016 to 2017 like the trend of squared returns. But the broken line can not accurately describe tomorrow's variance well. High squared returns will be accompanied by a high future variance. But the trend of predicted variance will fall sharply after the highest phase i.e. the red line segment which dropped at the end of 2016 in the diagram. This phenomenon is unrealistic.

Moving average variance shows also the non-normality of portfolio returns. The tails of distribution by dynamic variance are fat.[40] Here let us combined with the formula to see: Higher return brings relevant higher tomorrow's variance by 1/m times of which the opposite is also right. So the determination of m will affect the value of σ^2_{t+1} and it will further affect σ_{t+1}. A smooth trend of σ_{t+1} because of a large number of m observations.[41] There is a strict correlation between m and σ_{t+1} but hard to determine the value of m in practice.

[40] Christoffersen(2012), S. 143

[41] Christoffersen(2012), S. 69

Figure 3: Moving average variance on past 10 objects



So a more accurate tool for estimation of dynamic variance is needed. And it is necessary for the analysis of distribution for returns.

The following figure 4 shows GARCH(1,1) forecast variance under normal distribution from January 2015 to January 2020. According to results of the GARCH(1,1) volatility model(i.e. log-likelihood shows values of parameter mu, omega, alpha1 and beta1). Now note r_t as series of log returns. The actual σ_t^2 depends on beta equal to 0.5523 and alpha equal to 0.1075. The independence parameter omega is 0.0037. So GARCH variance model can be written by:

$r_t = 0.017 + a_t, \ \ a_t \sim N(0,1)$

$\sigma_t^2 = 0.0037 + 0.1075 a_{t-1}^2 + 0.5523 \sigma_{t-1}^2$

Figure 5 is t distributed GARCH(1,1) volatility model for Alibaba's log returns.The actual σ_t^2 depends on parameter alpha and beta, which are both equal to (1.0e-08). Omega is the constant parameter of the GARCH model and equal to 0.0011. The following formula is given by results in console(in the Appendix):

$r_t = 0.001 + a_t, \ \ a_t \sim \tilde{t}(d)$

$\sigma_t^2 = 0.0011 + (1.0e-08) a_{t-1}^2 + (1.0e-08) \sigma_{t-1}^2$

Through graphs, we can see the GARCH volatility model is more accurate to forecasting tomorrow's variance and especially at the extreme value of stock returns. And t distributed GARCH(1,1) volatility model describes dynamic variance is more detailed than normal distributed GARCH(1,1). That also means, the student t distribution can capture the change of variance due to change of stock returns better than normal distribution. The GARCH variance model proved also Non-normality distribution of stock returns because the distribution function of returns depends on all past variance.[42]

Figure 4: Normal distributed GARCH(1,1) volatility model for Alibaba's returns

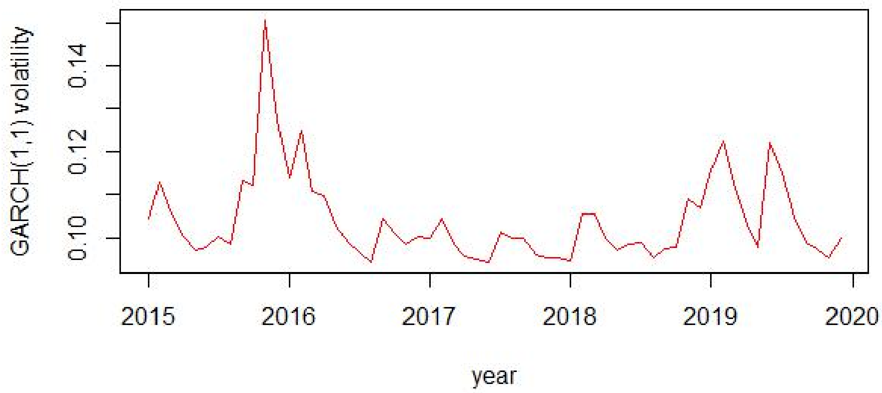
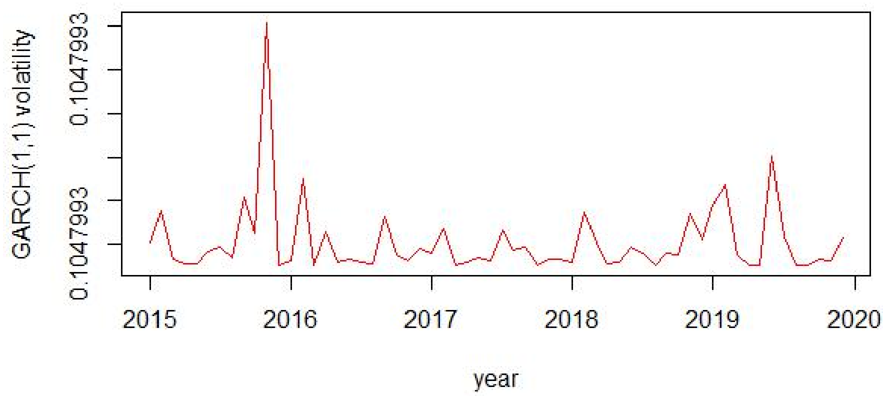


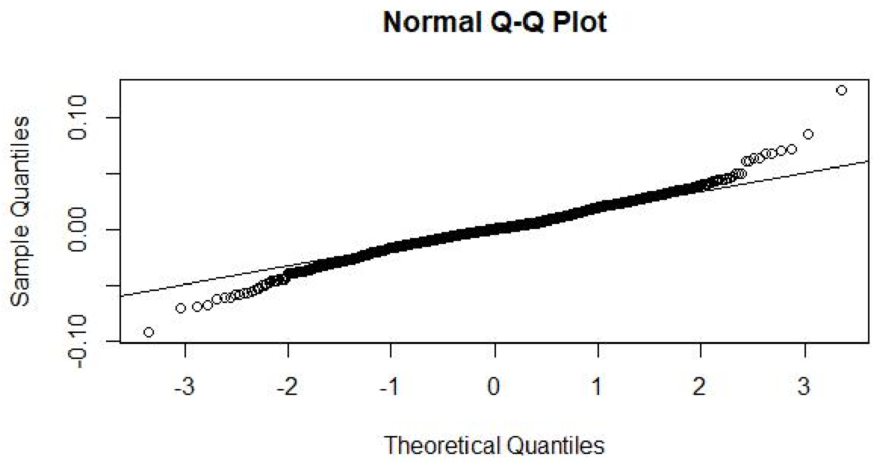
Figure 5: Student t distributed GARCH(1,1) volatility model for Alibaba's returns



[42] Christoffersen(2012),S. 121

Figure 6 is empirical quantiles of stock returns against normal distribution. The left and right tails are obviously diverged from the straight line. This QQ plot of returns shows the feature of fat tails is intuitive. In the univariate models, the daily Alibaba Group stock returns do not conform to normal distribution, the result is same as figure 1(Histogram of daily Alibaba log returns and normal distribution). So that the QQ plot confirms stock return distributed under non-normality.

Figure 6: QQ plot of daily returns for Alibaba Group from January 2015 to January 2020

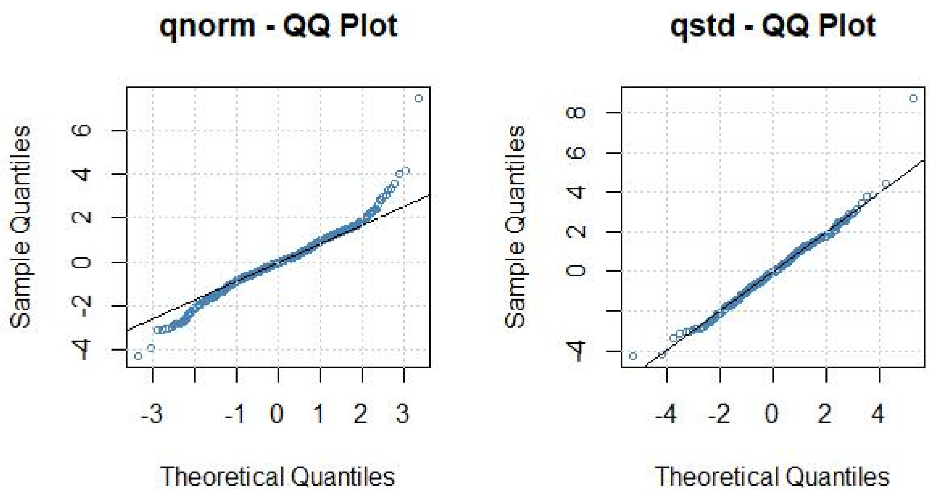


Use QQ plot and GARCH model together to show goodness-of-fit for asset returns. Figure 7 is a comparison for QQ plot of daily stock returns with normal distributed GARCH model and t distributed GARCH model. The left pattern is GARCH stock returns against normal distribution. The right one is GARCH stock returns against student t distribution.

It is clear to find that the right plot seems to fit the straight line better than the left graph, even though there is still some extreme value deviating from the straight line. There are fat tails on left and right sides of normally distributed GARCH shocks. Using the result obtained here: The student t distribution of returns is more consistent with sample data, so it is a better alternative approach than normal distribution in the univariate case.

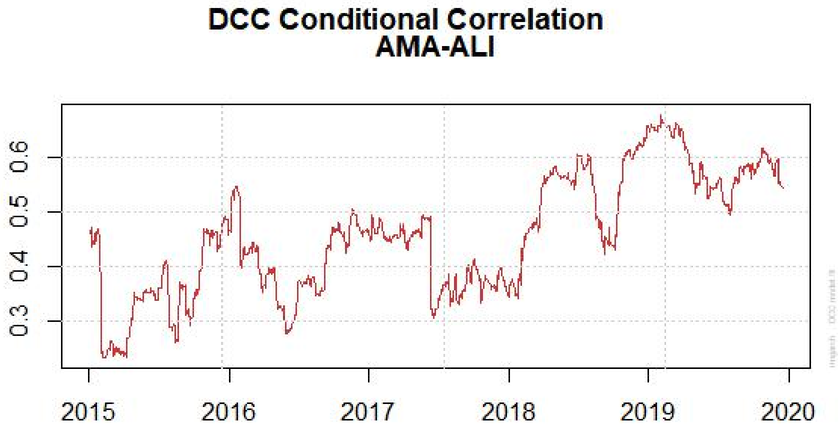
Furthermore, QQ plot for GARCH shocks against skewed t distribution can also be built. The plot will be shown in Appendix because the result is not significantly different from the t distribution.

Figure 7: QQ plots of daily stock returns with normal distribution and student t distribution using GARCH(1,1)



Then plot the dynamic conditional correlation for Alibaba Group daily log returns and Amazon log returns. It is clear to see that those two stock returns have positive correlation property. And the trend of conditional correlations fluctuates over time. Overall the trend of the two companies, their conditional correlation is growing. In 2019, the correlation reached a higher point.

Figure 8: Conditional correlations for Alibaba Group index and Amazon index with DCC model and GARCH model.

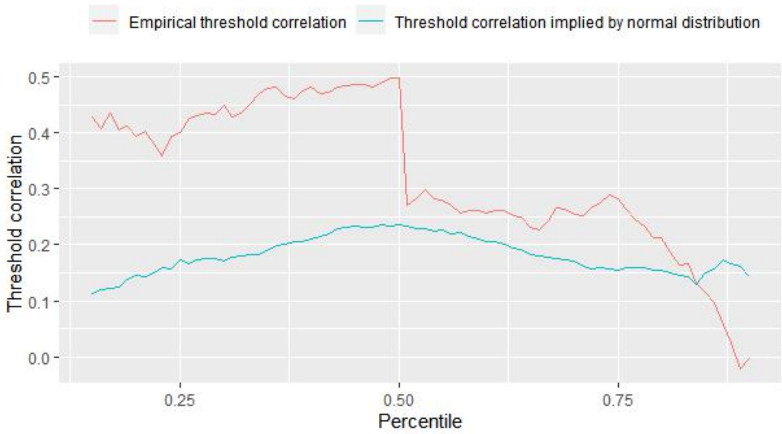


The red line is the threshold correlation with daily log returns from Alibaba Group and Amazon. And the blue lines describe threshold correlation from normal distribution with correlation matching those asset returns.

These two lines are obviously not matched with each other, even though there exist points where two lines intersect. The threshold correlation from two asset returns is higher than the threshold correlation by normal distribution with corresponding correlation as those returns overall. The blue line shows that the normal distribution between Alibaba's returns and Amazon's returns is asymmetric. And the threshold correlation will be higher when there are large positive returns on the left side of figure 9. The large positive asset returns have much higher correlation than large negative returns.

The result against the assumption once again. The multivariate threshold correlation from Alibaba index and Amazon index is non-normal distributed.

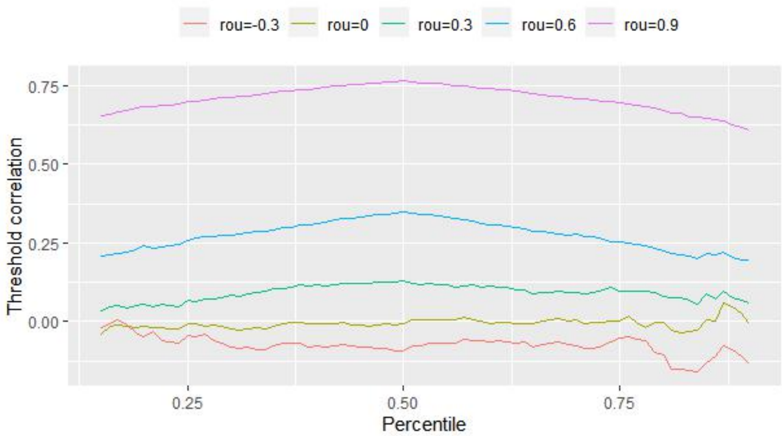
Figure 9: Threshold correlation plot for Alibaba Group versus Amazon and threshold correlation from normal distribution with correlation matching sample data.



The following figure 10 is a threshold correlation from bivariate normal distribution with different values of correlation ρ versus percentile. These lines proved again that threshold correlation from multivariate normal distribution is asymmetric.

For large correlations, the threshold correlation will be changed accordingly. For some particular correlation value, for example, when ρ equal to zero, the threshold correlation can be zero too. So the multivariate normal distribution model is not fit to large asset returns with extreme probability value, which has a large threshold correlation.[43]

Figure 10: Threshold correlation from bivariate normal distribution

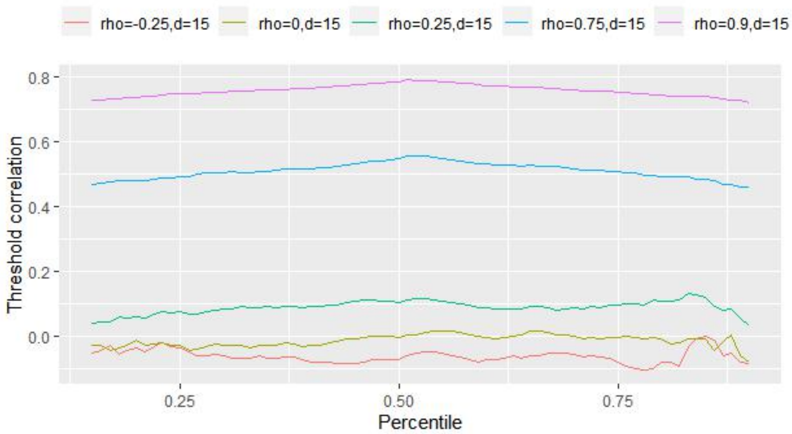


Now considering threshold correlation of multivariate t distribution. It is clear to see

[43] Christoffersen(2012),S. 197

more flexible degrees of tail dependence using bivariate t distribution of asset returns. Comparing the broken line in yellow on Figure 10 and 11, when correlation ρ equal to zero, the threshold correlation by multivariate t distribution is not going to zero. The multivariate t distribution describes more accurate than normal distribution for asset returns, which have large threshold correlations.[44]

Figure 11: Threshold correlation implied by bivariate t distribution



The multivariate normal distribution can not describe threshold correlation of daily asset returns well. The multivariate t distribution can be used to describe the larger threshold correlation but hard to calculate because conditions of parameter d is restrict.[45] Therefore, considering copula models, which can combine all univariate functions into a multivariate distribution model and simulate the requisite parameter easier.

When these two copula models using the same tau as the correlation parameter, the flagrant contrast is obvious. But the same character is the probability density of normal copula and t copula are both symmetric.

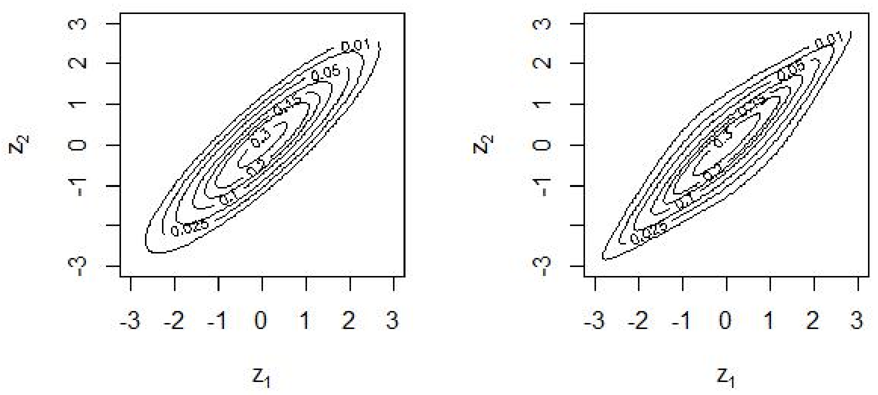
The outer contour has a lower value of probability with extreme variables for the combination of assets. But the shape of contour probability for normal copula is elliptical, which means this copula is not conformed to large(whatever positive or negative) sample data. However, consider the bottom-left and top-right corner of the

[44] Christoffersen(2012),S. 203

[45] Christoffersen(2012), S. 203

contour plot for t copula. The probability contour plot for t copula looks almost like a symmetrical quadrilateral. Extreme variables will be considered better in t copula model.[46]

Figure 12: Contour probability plots for normal and student t copula



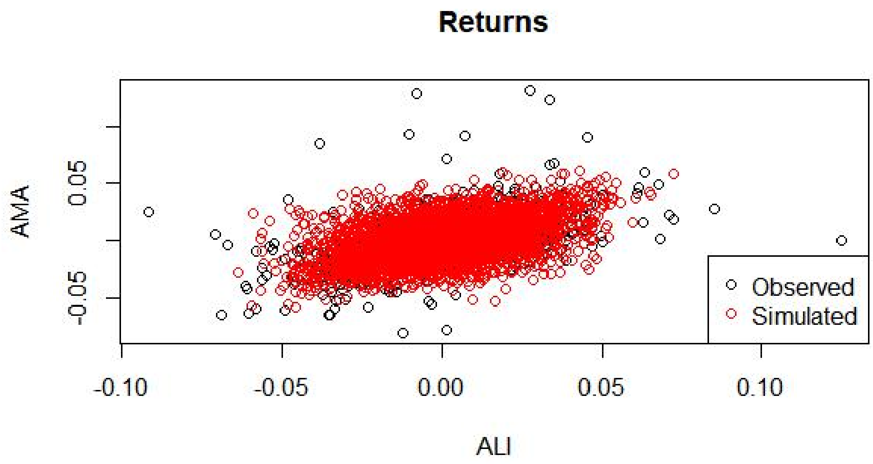
Finally, apply sample data i.e. Alibaba's and Amazon's asset returns to the copula model. From the above probability contour for normal copula and t copula, we can get a conclusion is that t copula describes sample data is more accurately. So that constructing scatter points plot for sample data and simulated objects under normal marginals and t copula model.

The points in black are sample data. And red points describe simulated observations, which conform to t copula with normal marginals function. Almost all of the simulated observations are fit to sample data, even though there are regardless of some extreme value. Some black points are outside with no red points.

Therefore, the t copula model gives the way to link marginal distributions into a joint distribution model for asset returns. And it describes distributions of asset returns well.

[46] Christoffersen(2012), S. 210

Figure 13: Scatter chart for Alibaba's returns and Amazon's returns under student t copula



5 Conclusion

Now return to the original questions at the beginning of this paper. How to form the proper joint distribution of asset returns? There are three approaches to constructing the joint distribution of asset returns. The best of those approaches is student t copula model. And what are the criteria for a good approach to constructing joint distribution of asset returns? The univariate model and Monte Carlo simulation are the most important criteria for evaluation.

The multivariate normal distribution is convenient for computation because linear combination of normal variables is also normally distributed in the multivariate case, which means calculation of modeling is simple[47] But this model can not accurately describe features of asset returns. It gives an inflexible way to explain the shape of tails. And asset returns do not conform to normal distribution in fact. If constructing joint distribution of returns data using multivariate normal distribution model, it will create exorbitant price and underestimates the probability of extreme returns.[48] On the other hand, it increases risk factors for investors or for risk managers.

[47] Christoffersen(2012),S. 198

[48] Christoffersen(2012),S. 193

The multivariate t distribution of asset returns is flexible for large threshold correlations. But this model is hard to calculate because of the restrict condition of parameter d.[49]

The copulas can combine all those details from kind of distributions and used to link individual marginal distributions and returns to form a proper multivariate distribution model. The normal copula is better than multivariate normal distribution, but it doesn't have enough dependence for different tails of distributions.[50] The t copula model allows large threshold correlation with extreme variables. Or rather, the most important advantage of the t copula model is the inclusiveness. That means t copula allows many factors(e.g. value of mean, deviations, kurtosis and skewness for a density function, etc) together to obtain a joint distribution for asset returns. This approach based on the univariate model of individual assets and linked them together to produce a multivariate risk model of asset returns.

The univariate model is used to measure individual portfolio return and forecast time-varying variance. Besides, the non-normality between the distribution of each asset can be captured. Then constructing multivariate distribution of asset returns based on the univariate model and simulation method i.e. Monte Carlo.[51]

The t copula for the joint distribution of asset returns shows also good-fitness to integrated risk management.

---

[49] Christoffersen(2012),S. 203

[50] Christoffersen(2012),S. 207

[51] Christoffersen(2012),S. 17

Appendix

References

Master's thesis statement of originality

I hereby confirm that I have written the accompanying thesis by myself, without contributions from any sources other than those cited in the text and acknowledgements. This applies also to all graphics, drawings, maps and images included in the thesis.

Place and date   Signature