Dependence Modeling The Rate of Returns of Two Stocks(NVDA, TSM): A Copula Approach

Ray Li

Supervisor: Dr. Orla Murphy

Department of Mathematics and Statistics, Dalhousie University

November 2024

Submitted in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science:

Combined Honours in Statistics and Economics

Acknowledgements

This research topic emerged from my keen interest in both statistics and economics. I would like to thank my supervisor Dr. Orla Murphy for her unwavering support, guidance, and encouragement throughout the development of this thesis. She not only guided me in learning about copula models, but also gave me valuable advice and patience when I encountered difficulties with various concepts. It is because of her continuous support that I was able to complete my thesis. Her expertise and insightful feedback were invaluable and greatly enhanced the quality of my work. In addition, I would also like to thank my family and friends for their support and encouragement.

Table of Contents

1 Introduction

Investors utilize hedging strategies to balance risk and return, aiming to minimize potential losses while maximizing gains. A critical aspect of effective hedging and portfolio optimization is understanding the co-movement of stock returns—that is, the dependencies between returns. However, modeling financial dependencies poses significant challenges due to the complex and dynamic nature of financial markets. In the past 5 years, Nvidia Corporation (NVDA) has experienced significant growth in its stock price and, notably, does not produce its own chips; instead, it relies on Taiwan Semiconductor Manufacturing Company Limited (TSM) for chip manufacturing (Yoon, 2023). The intricate relationship between these two firms suggests a potential dependence in their stock returns. By investigating this dependence through copula models, we aim to enhance hedging strategies and improve portfolio management involving these stocks.

In the field of economics, Embrechts is one of the first economists to use copula in 1999, to solve the problem of insurance companies suffering from property losses due to multiple major disasters (Embrechts et al, 1999). Many researchers analyzed the dependence structure of market indices (Peng et al, 2012), risk level (Embrechts, 1999), exchange rate (Du et al 2017), etc. But less people have studied the dependence between two stocks.

This research analyzes the dependence structure between the stock returns of two firms, Nvidia (NVDA) and Taiwan Semiconductor Manufacturing Company (TSMC), using copula models. The specific objectives are to assess the presence and nature of tail dependence between the two stocks, model the non-linear and tail dependencies using various copula functions and evaluate the effectiveness of chosen model.

2 Background

2.1 Time Series Model

Before diving into correlations and copula models, we need to discuss the property of the data. Our raw data are the daily stock prices and are highly autocorrelated (dependence on their own past values). These patterns may distort how two variables, such as stock returns, depend on each other. This makes it more difficult to identify and examine their actual simultaneous relationship. Also, copula models assumed the input data (marginals) are independent and identically distributed. If time-dependent patterns remain, the data may not satisfy these assumptions, leading to mistaken results. Therefore, we need to eliminate the effect of time using a time series model so that the residuals will be independent and have constant variance if there is a good fit (Engle, 1982). This allows the analysis to focus on the instantaneous dependence (co-movements) between the two variables, which is the primary interest when using copula models. The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, introduced by Bollerslev (1986), is an extension of the Autoregressive Conditional Heteroskedasticity (ARCH) model initially developed by Engle (1982). These models are widely used in finance and econometrics to model time-varying volatility in time series data, which is particularly common in financial returns.

We will only focus on the $GARCH(1,1)$ model that we will use in this paper, which is also the simplest and most widely used specification of the GARCH family (Hansen and Lunde, 2005). The mathematical form is:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2,$$

where:

σ_t^2: Conditional variance.

$\omega > 0$: Constant term.

α: Impact of recent shocks on current volatility.

β: Persistence of past volatility.

ε_{t-1}^2: Previous period's squared residual.

σ_{t-1}^2: Previous period's variance.

Next, we find the residuals by the fitted $GARCH(1,1)$ model. However, we need to be careful before using those residuals because they still contain volatility information (Bollerslev, 1986). Therefore, standardized residuals $z_t = \frac{e_t}{\sigma_t}$ are obtained by dividing the raw residuals (e_t) by the conditional standard deviation (σ_t) estimated from the $GARCH(1,1)$ model. Through the standardizing process, we can remove the time-varying volatility effect and make residuals, ideally, behave like white noise which are independent of time (Engle, 1982). Doing so can help us distinguish if the dependency is because of the time effect on both variables. These residuals will be explored in more depth in the exploratory analysis to evaluate the effectiveness of the GARCH model and to verify their appropriateness for copula modeling.

2.2 Associations

Traditional methods for measuring dependence between financial assets, such as the Pearson correlation coefficient, have significant limitations in capturing the true complexity of financial market relationships. The Pearson correlation assumes a linear relationship and measures only the strength and direction of a linear association between two variables, also known as bivariate correlation. It is defined as:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y},$$

where $cov(X, Y) = E(XY) - E(X)E(Y)$ is the covariance of X and Y, and σ_X and σ_Y are their respective standard deviations. Pearson's correlation has a range from -1 to 1. If there exists a perfect positive linear relationship between two variables, the correlation equals to 1; if there is no linear relationship, the correlation equals to 0; if there is a perfect negative linear relationship, the correlation equals to -1 (Weisstein. Eric W, 2024).

However, financial asset returns often exhibit non-linear, asymmetric, and tail-dependent behaviors, especially during periods of market shocks (Embrechts et al, 2002). These non-linear dependencies mean that extreme movements in one asset may not be adequately reflected in the correlation measure with another asset. Consequently, relying solely on Pearson correlation can lead to an incomplete or misleading understanding of the dependence structure between assets. This limitation underscores the need for models that can capture non-linear dependencies and tail dependence, providing a more accurate and nuanced representation of how asset returns co-move under various market conditions. To address these limitations, we could have used the rank-based alternatives: Kendall's tau (τ) and spearman's rho (ρ_s). These are rank-based measures employed to better capture non-linear dependencies.

Kendall's tau (τ), commonly referred to as Kendall rank correlation coefficient, is a measure of rank correlation: the similarity of the orderings of the data when ranked by each of the quantities (9. Kendall, 1938). Mathematically:

$$\tau = 1 - \frac{2(number\ of\ discordant\ pairs)}{\binom{n}{2}},$$

where n is sample size, and $\binom{n}{2} = \frac{n(n-1)}{2}$ (Nelsen, 2001).

Spearman’s rho (ρ_S), also referred as Spearman's rank correlation coefficient is defined as the Pearson correlation coefficient between the rank variables (Spearman, 1904). Spearman correlation between two variables is equal to the Pearson correlation between the rank values of those two variables. While Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships (whether linear or not). Mathematically:

$$\rho_S = \frac{Cov[Ran(X), Ran(Y)]}{\sigma_{Ran(X)}\sigma_{Ran(Y)}},$$

here $Ran(X)$ and $Ran(Y)$ are rank variables corresponding to X and Y (Myers et al, 2003).

These measures are advantageous because they do not assume a linear functional form and are less sensitive to extreme values, which is crucial for analyzing the often-volatile nature of financial returns. While Kendall's τ and Spearman's ρ_S provide useful insights into rank correlations, they do not fully describe the joint distribution of random variables. This is why Sklar's theorem and copulas are valuable to analyze the financial data.

2.3 Copulas

Copula models offer a flexible framework for modeling the joint distribution of multiple financial assets (Embrechts et al, 2002). A copula is a mathematical function that links univariate marginal distribution functions to form a multivariate distribution, thereby allowing the separation of marginal behaviors from the dependence structure between variables (Nelson, 2006). One of the foundational principles underpinning copula theory is Sklar's theorem (Sklar, 1959), which states that every multivariate cumulative distribution function

$$H(x_1, \dots, x_d) = \Pr\,[X_1 \le\ x_1, \dots, X_d \le\ x_d]$$

of a continuous random vector $(X_1, X_2, \dots, X_d)$ can be expressed in terms of its marginals $F_i(x_i) = \Pr\,[X_i \le x_i]$ and a unique copula *C*:

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$$

Similarly, given any copula *C* and marginal distribution functions (F_X and F_Y), a valid joint distribution $F_{X,Y}(x, y)$ can be constructed as $C\big(F_X(x), F_Y(y)\big)$. Also, $C :$ $[0,1]^d \to [0,1]$ is a *d*-dimensional copula if *C* is a joint CDF (cumulative distribution function) of a *d*-dimensional random vector on the *d*-dimension unit space $[0,1]^d$ with uniform marginals (Nelson, 2006).

Building on the theoretical foundation, the following paragraphs will introduce and explore five specific bivariate copula models which will be analyzed, one of which will be chosen to measure the dependence structure of stock returns in this thesis. Each of the five specific bivariate copula models offers unique capabilities in capturing different aspects of dependence.

A special case outside of those five copulas is probably the simplest bivariate copula: The 2-dimentional **Independence copula**, which is a non-parametric copula of 2 independent $Uniform(0,1)$ distributed random variables,

$$C_\pi(u, v) = uv,$$

where u and $v\ \in (0,1)$ (Ruppert, 2011).

Before diving into other copulas, we need to define tail dependency to showcase the properties of those models. The tail dependency is showcased by tail dependence coefficient, which measures the likelihood of extreme co-movements between two variables in the upper tail (extreme highs) or the lower tail (extreme lows) of their joint distribution. Mathematically, for two random variable X and Y, the upper tail dependence coefficient (λ_u) and lower tail dependence coefficient (λ_l) are defined as below:

- $\lambda_u = \lim_{q \to 1} P(Y \leq F_Y^{-1}(q) | X \leq F_X^{-1}(q))$,
- $\lambda_l = \lim_{q \to 0} P(Y > F_Y^{-1}(q) | X > F_X^{-1}(q))$,

where $F^{-1}(q) = \inf\{x \in R: F(x) \geq q\}$, which is the inverse of the CDF for q (McNeil et al, 2005). Thus, if the joint distribution has tail dependency, then λ_u or λ_l or both would be larger than zero, if not, then equal to zero. In addition, we call a joint distribution symmetric in tail dependence when $\lambda_u = \lambda_l > 0$, asymmetric in tail dependence when $\lambda_u \neq \lambda_l$, and no tail dependence when $\lambda_u = \lambda_l = 0$.

One of the most popular copulas in financial modeling due to its simplicity and mathematical tractability, is **Gaussian/Normal copula**, which is derived from the multivariate normal distribution. For two uniform random variables U and V, the Gaussian copula with correlation parameter θ is defined as (Nelson, 2006)

$$C_{Gaussian}(u, v;\ \theta) = \Phi_\theta\left(\Phi^{-1}(u), \Phi^{-1}(v)\right),$$

where:

- Φ^{-1} is the inverse of the cumulative distribution function (CDF) of standard normal distribution,
- Φ_θ is the bivariate normal CDF with correlation θ.

Properties:

- No tail dependence, $\lambda_u = \lambda_l = 0$.

Gaussian copula is ideal for scenarios where the relationship is primarily linear, especially when extreme co-movements (tail dependencies) are not a primary concern. It serves as a baseline model to compare against more flexible copulas that can capture non-linear and tail dependencies.

The **Student t-copula** extends the Gaussian copula by incorporating degrees of freedom v, allowing it to model tail dependencies and capture more extreme co-movements between variables. Mathematical definition: the joint distribution of (u, v) is the t-copula with correlation parameter θ and degrees of freedom ν $(\nu > 0)$ if

$$C_t(u, v;\ \rho, \nu) = t_{\theta,\nu}(t_\nu^{-1}(u), t_\nu^{-1}(v)),$$

where:

- t_ν^{-1} is the inverse CDF of the univariate Student's t-distribution with ν degrees of freedom.
- $t_{\theta,\nu}$ is the bivariate Student's t CDF with correlation θ and ν degrees of freedom (Embretchs et al, 2002).

Properties:

- Symmetric tail dependence: $\lambda_u = \lambda_l > 0$

This copula is suitable for financial applications where extreme co-movements (both gains and losses) are of interest. It provides a more realistic modeling of risk by accounting for tail dependencies.

These copulas (Gaussian and Student t) are derived from elliptical distributions (normal or t-distributions) and therefore belongs to elliptical copula family. They are used to describe dependencies particularly when the relationships are approximately linear or symmetric (Mai and Scherer, 2012).

On the other hand, the following three copulas belong to Archimedean copula family. These copulas use a generator function and can capture non-linear and asymmetric (tail) dependencies (De Matteis 2001).

The **Clayton copula** is particularly adept at modeling dependencies in the lower tail, making it useful for assessing the likelihood of simultaneous extreme losses. Mathematically, the joint distribution of (u, v) is the Clayton copula with dependence parameter θ $(\theta > 0)$ if (Cherubini et al, 2004):

$$C_{Clayton}(u, v;\ \theta) = [\max\,(u^{-\theta} + v^{-\theta} - 1, 0)]^{-\frac{1}{\theta}},$$

with properties:

- Asymmetric tail dependence: lower tail, $\lambda_l > 0$ and $\lambda_u = 0$, effectively models the co-occurrence of extreme losses.
- Monotonic dependence: Model's dependencies increase as θ increases.

This modal is ideal for situations where the primary concern is the joint occurrence of extreme negative events, such as simultaneous stock price drops, which is critical for insurance, risk management, and hedging strategies.

The **Gumble copula**, on the contrary, is designed to model dependencies in the upper tail, making it suitable for assessing the likelihood of simultaneous extreme gains. Mathematically, the joint distribution of (u, v) is the Gumbel copula with dependence parameter θ $(\theta > 0)$ if (De Matteis, 2001):

$$C_{Gumbel}(u, v;\ \theta) = \exp\left\{-\left((-\ln u)^{\theta} + (-\ln v)^{\theta}\right)^{\frac{1}{\theta}}\right\},$$

with properties:

- Asymmetric tail dependence: upper tail, $\lambda_u > 0$ and $\lambda_l = 0$, effectively models the co-occurrence of extreme gains.
- Monotonic dependence: Models increasing dependencies as θ increase.

This copula is good for scenarios where the joint occurrence of extreme positive events is of interest, such as simultaneous stock price surges, which can inform investment strategies aimed at capitalizing on upward market movements.

The **Frank copula** is known for its ability to model dependencies without exhibiting tail dependences. It captures moderate dependence across the entire range of data. While Gaussian and Frank copula both are symmetric and have no tail dependence, Frank copula does not require a linear correlation assumption, making it a good choice for various financial applications. Mathematical definition: the joint distribution of (u, v) is the Gumbel copula with dependence parameter θ $(\theta \neq 0)$ if (Ruppert, 2011)

$$C_{Frank}(u, v;\ \theta) = -\frac{1}{\theta}\ln(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}),$$

with properties:

- No tail dependence: $\lambda_u = \lambda_l = 0$.
- Bounded dependence: Models increasing dependencies as θ increase.

Frank copula is particularly useful in scenarios where the dependency structure is moderate and symmetric without significant tail dependencies. It serves as a complementary model to copulas like Clayton and Gumbel by providing flexibility in capturing dependencies that are not extreme but still significant.

Table 1 summarizes the five copula models mentioned above, with the range of parameters and the independent case. Figure 1 provides simulated samples of 6 copula models (including independent case), where $\theta_{Gaussian} = 0.7$, $\theta_t = 0.6$ and $v = 6$, $\theta_{Clayton} = 2$, $\theta_{Gumbel} = 2$, and $\theta_{Frank} = 5$.

Table 1: Few bivariate copulas.

Name	Distribution function	Parameter	Independence
Gaussian	$C(u,v) = \Phi_\theta(\Phi^{-1}(u), \Phi^{-1}(v))$	$\theta \in [-1,1]$	$\theta = 0$
t-copula	$C(u,v) = t_{\theta,v}(t_v^{-1}(u), t_v^{-1}(v))$	$\theta \in [-1,1]$, $v > 0$	$\theta = 0$
Clayton	$C(u,v) = [\max(u^{-\theta} + v^{-\theta} - 1, 0)]^{-\frac{1}{\theta}}$	$\theta \in (0,\infty)$	$\theta = 0$
Gumbel	$C(u,v) = \exp\{-\left((-\ln u)^\theta + (-\ln v)^\theta\right)^{\frac{1}{\theta}}\}$	$\theta \in [1,\infty)$	$\theta = 1$
Frank	$C(u,v) = -\frac{1}{\theta}\ln(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1})$	$\theta \in R, \theta \neq 0$	$\theta \to 0$

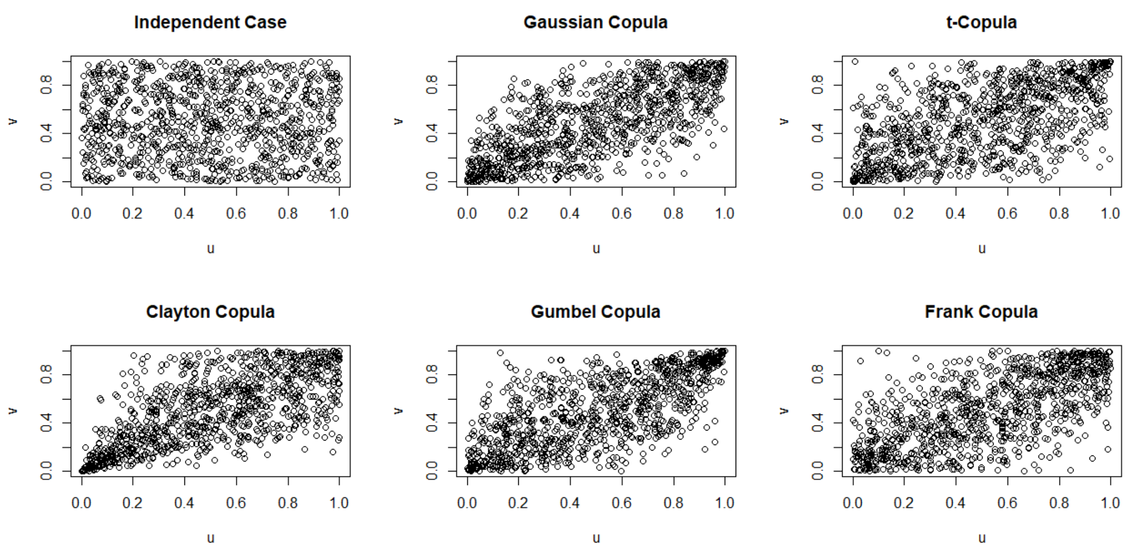

Figure 1: Scatter plots of 1000 random numbers generated from the proposed bivariate copulas.

In addition, there also exists other type of **bivariate asymmetric copulas**, which are different from asymmetry in the tail dependence. A copula C is symmetric or exchangeable if $C(u,v) = C(v,u)$ for all $u, v \in [0,1]$, where u and v is the uniform margins, otherwise asymmetric or non-exchangeable. Therefore, asymmetric copula $C(u, v) \neq C(v, u)$. For instance, the Frank & Gumbel copula is a combination of Frank and Gumbel copulas which shows asymmetry (Mukherjee et al, 2018).

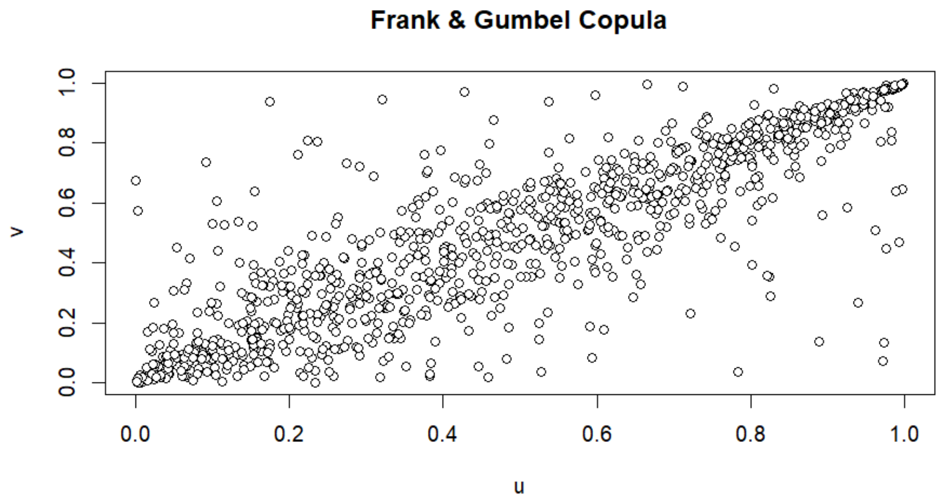

Figure 2: Scatter plot of random numbers generated from the Frank & Gumbel copula (asymmetric).

However, we will only focus on those widely used copula models such as Gaussian, t-Copula, Gumbel, Clayton, and Frank copulas to analyze the dependence structure between NVDA and TSM stock returns. The combined or mixed copula models, such as the "Frank & Gumbel" copula, are not considered because they introduce additional complexity which will not be discussed.

3 Data

3.1 Description

The financial data for Nvidia (NVDA) and Taiwan Semiconductor Manufacturing Company Limited (TSM) were sourced from Yahoo Finance, a reputable and widely used platform that provides historical stock prices and financial information. The dataset consists of daily closing prices for both companies over the past five years, from Sep 24th, 2019, to Sep 23rd, 2024. Using daily data allows for a large number of observations (1,258) to perform reliable statistical analysis.

The initial variables include the date and daily closing price for both firms represented as NVDA for Nvidia and TSM for Taiwan Semiconductor Manufacturing Company. We generated a new column titled "Time" that contains only natural numbers from 1 to 1,259 as time index to replace the categorical data with the numeric data. The steps involved in preparation after downloading the data include:

- Data cleaning: checked for missing values, removed dividend rows, and aligned the data and time of two variables.
- Rate of returns calculation: computed daily rate of returns for each stock using the formula: $\frac{P_t - P_{t-1}}{P_{t-1}}$, where P_t is the closing price at time t and P_{t-1} is the closing price at time $t-1$.
- The log returns are defined as ln $(\frac{P_t}{P_{t-1}})$.

We focus on rate of returns instead of the raw price because the rate provides a percentage change of value, which allow us to make comparisons between different stocks that have various pricing scales. For example, an increase of $1 on a $100 stock versus a $10 stock is very different. Also, rate of returns directly measure the profitability of an investment, which is a key metric for investors.

3.2 Visualizations

We generated plots of the time series for the original stock (closing) price:

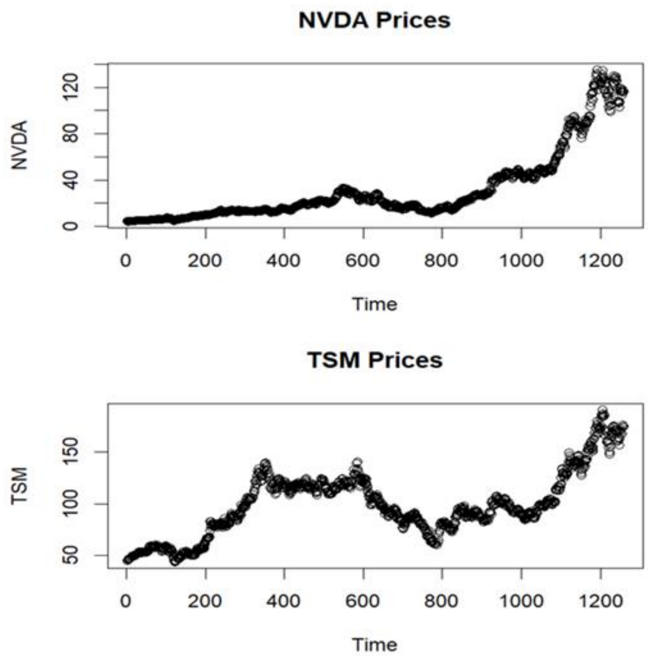

Figure 3: Time series plots of NVDA and TSM stock (closing) prices.

And rate of returns for both firms:

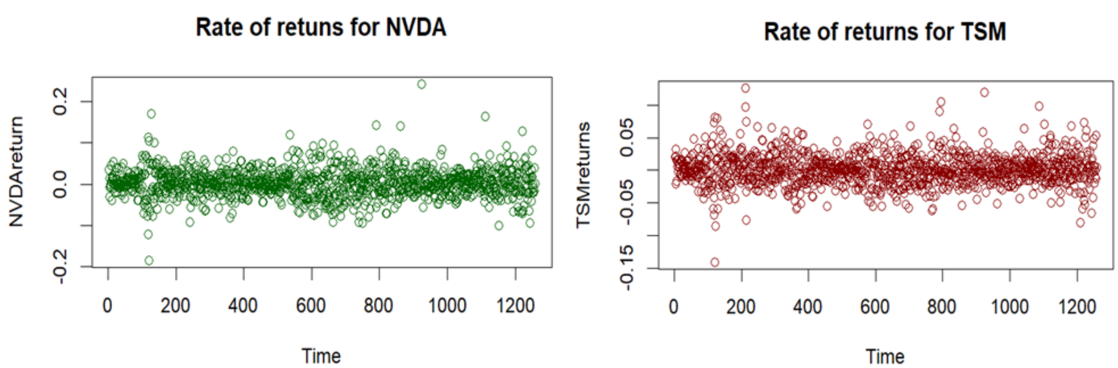

Figure 4: Time series plots of NVDA and TSM rate of returns.

These two figures above show the non-stationary property of stock prices, where they have a natural upward trend over time. The rate of returns is stationary since they fluctuate around the mean and do not have obvious long-term trends.

Additionally, Figure 5 below shows the rate of return plot and log-return plot with respect to the relations between those two firms.

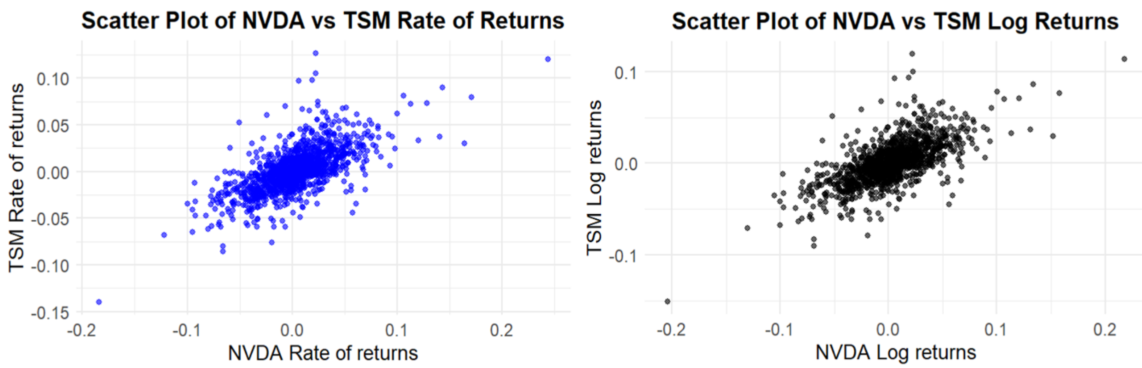

Figure 5: Scatter plots for rate of returns (left) and log-returns (right).

There is no obvious difference between the plot of rate of returns and log returns. This can also be shown through the histograms of rate of returns and log returns (Appendix, Figure 6 and 7). The histograms show an approximately bell-shaped curve with some outliers; however, they seem be more spread out than normal distribution, therefore we might need further examination for normalities. Overall, for simplicity, we will only focus on the rate of returns.

3.3 Statistical Summaries

Table 3: Statistical summary of Nvidia (NVDA) rate of returns.

Summary of NVDA Returns					
Min	1st Quartile	Median	Mean	3rd Quartile	Max
-0.1844	-0.0155	0.0031	0.0032	0.0221	0.2436

Table 4: Statistical summary of Taiwan Semiconductor Manufacturing Company Limited (TSM) rate of returns.

Summary of TSM Returns					
Min	1st Quartile	Median	Mean	3rd Quartile	Max
-0.1403	-0.0122	0.0005	0.0014	0.0139	0.1265

The statistical summaries contain mean, median, 1st and 3rd quartile, and maximum and minimum values. Most returns fall within the range between -20% to 20%, and the mean for NVDA is approximately 0.32%, suggesting a slight overall upward trend among the period analyzed. On the other hand, TSM's mean rate of return is 0.14%, indicating more moderate growth because it has a higher start price. Both stocks exhibit considerable variability, as reflected by their respective standard deviations of 3.38% (NVDA) and 2.37% (TSM).

3.4 Association Analysis

The association analysis aims to understand the linear and non-linear dependencies between NVDA and TSM returns. Understanding associations provides insights into how closely the returns of these two companies move together, which is crucial for assessing co-movement risk and developing effective hedging strategies. The R code is included in Appendix 6.3.

Pearson’s correlation coefficient of the stock price of NVDA and TSM is 0.7754 which is a positive correlation, indicating a mod-strong linear relationship between the two stock prices, implying that their daily closing prices move together quite significantly, i.e. when NVDA's stock price increases, TSM's stock price also tends to increase, and vice versa. The high positive correlation of this magnitude typically indicates that both stocks are highly affected by similar market conditions or industry trends. This makes sense as Nvidia relies on Taiwan Semiconductor Manufacturing Company to produce chips, as we mentioned in the introduction. In addition, the Pearson correlation between NVDA and TSM’s rate of returns is 0.6685, still indicating a positive relationship, although a bit weaker than the correlation between their stock prices. Since the rate of returns focus on percentage changes rather than absolute values, their correlation is less influenced by the shared long-term trends.

The Spearman's rho of returns is 0.6662 and the Kendall's tau is 0.4870. Spearman's rho of NVDA and TSM rate of returns shows a relatively strong monotonic and positive relationship. And the Kendall's tau emphasizes the consistency of the co-movement across observations. The value (0.478) indicated that roughly half of the pairs of observations are concordant (both move in the same direction), reinforcing the tendency of the two variables to increase or decrease together, ad maintaining the possibilities of some inconsistencies.

4 Results

4.1 Residuals

We used a GARCH(1,1) model to obtain standardized residuals from the rate of returns of each of the two companies. We need to examine whether this time series model is appropriate for our data. Thus, we visually inspect the residuals for any remaining autocorrelation after fitting a GARCH(1,1) model, and no lags appear significant. Therefore, the model is indeed capturing the daily dynamics appropriately (Hansen and Lunde, 2005).

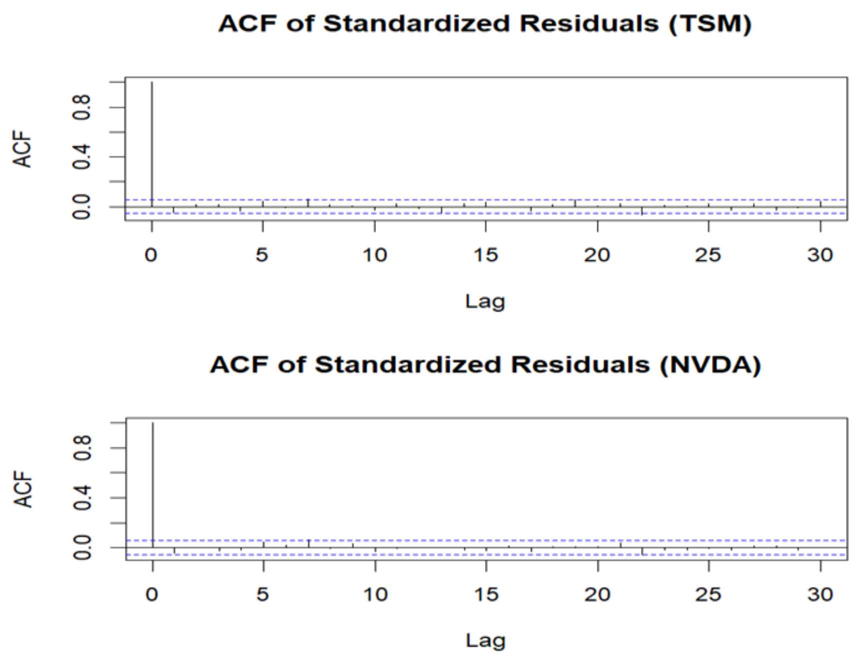

Figure 8: ACF plots of standardized residuals after fitting $GARCH(1,1)$ model.

In addition, we run the Ljung-Box test which is a formal statistical test used to determine whether any significant autocorrelation remains in the residuals after fitting a model (Bollerslev, 1986). The corresponding p-values are 0.6944 and 0.1391 for NVDA and TSM standardized residuals, respectively. There is insufficient evidence that the null hypothesis of "no autocorrelation" can be rejected, implying that the residuals are independent, and the model has good fit. A statistical summary of standardized residuals can be found in Appendix 6.2, Table 5.

The histograms in Figure 9 show that both NVDA and TSM residuals exhibit a distribution centered around the 0, and some extreme values on both the positive and negative ends suggests occasional outliers, consistent with the heavy-tailed nature of financial data.

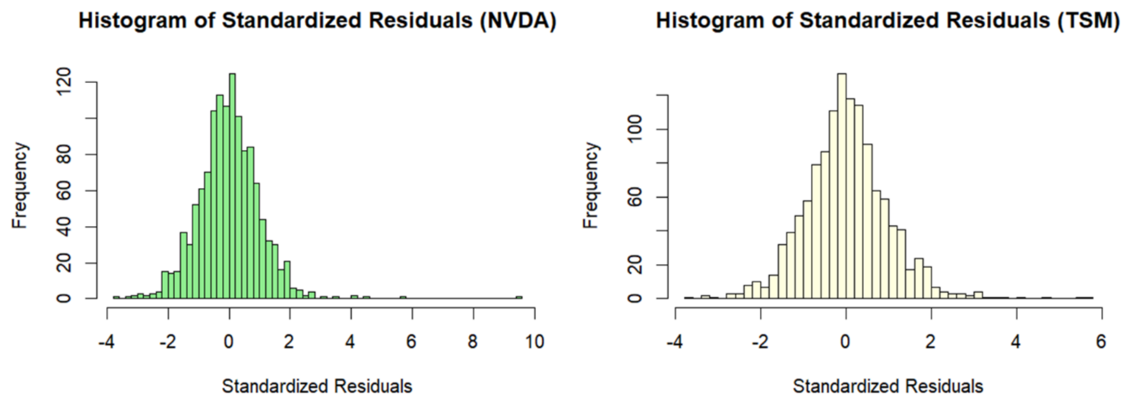

Figure 9: Histograms of standardized residuals of NVDA (left) and TSM (right).

4.2 Copula model selection

Since the tail dependence coefficients can provide an empirical basis for choosing among candidate copula models (Patton, 2006). To determine an appropriate copula model for the standardized residuals, we calculate the tail dependencies of the standardized residuals with a preset threshold and compare them with those various copula models (mentioned in section 2.3). We set the threshold to be 0.1 and 0.9 as lower and upper quantiles, respectively, yielding 46 observations above the upper quantile and 64 observations below the lower quantile. The upper tail dependence coefficient (λ_u) equals 0.0366 and lower tail dependence coefficient (λ_l) equals 0.0509, which shows statistically significant (p-value < 0.0001) weak tail dependence in the permutation test (Appendix 6.3). These results indicate weak but slightly stronger co-movement during extreme negative returns.

By the properties (mentioned in section 2.3) of those bivariate copula models in Table 1, the Clayton copula would be able to measure lower tail dependence, and the

Student t-copula which measures both tail dependences. Thus, these two copula models may be appropriate for our data.

Therefore, we run the parametric bootstrap-based goodness-of-fit test for Clayton copula and Student t-copula. Our null hypothesis is that the specified copula model is a sufficient fit for the observed data. The result shows that for t-copula, the first parameter (θ) is 0.67, the second parameter (v) is 6.33, and p-value equals 0.1084. Thus, there is insufficient evidence to reject the null hypothesis for t-copula. On the other hand, the test results of Clayton copula show that parameter (θ) is 1.86, and p-value equals 0.0005. Therefore, the results indicate sufficient evidence to reject the null hypothesis for Clayton copula (Appendix 6.3). Overall, t-copula is more sufficient to use than Clayton copula for standardized residuals.

We can find the same conclusion through comparing the rank plots of our observed data (Figure 11) with simulated plots of Clayton copula and Student t-copula (Figure 12). In Figure 11, we transformed the standardized residuals into a uniform distribution with the range of [0,1], by ranking them and dividing their rank with the number of residuals. In Figure 12, we generate simulations of two-dimensional Student t-copula and Clayton copula with 1000 points and unknown parameters (Appendix 6.3).

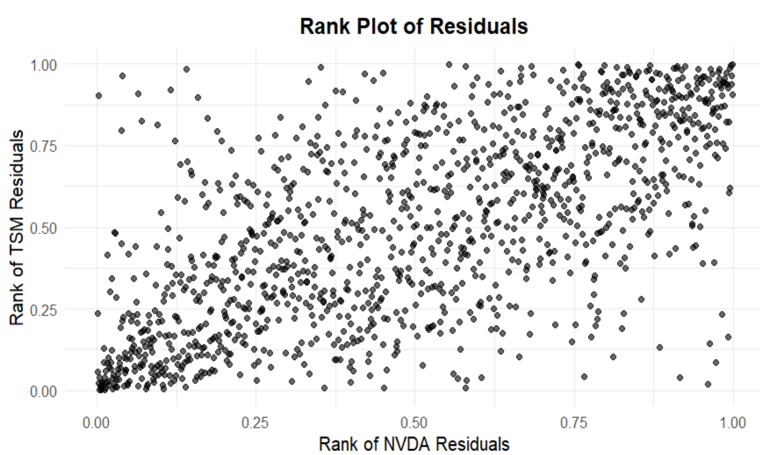

Figure 11: Rank plot of standardized residuals

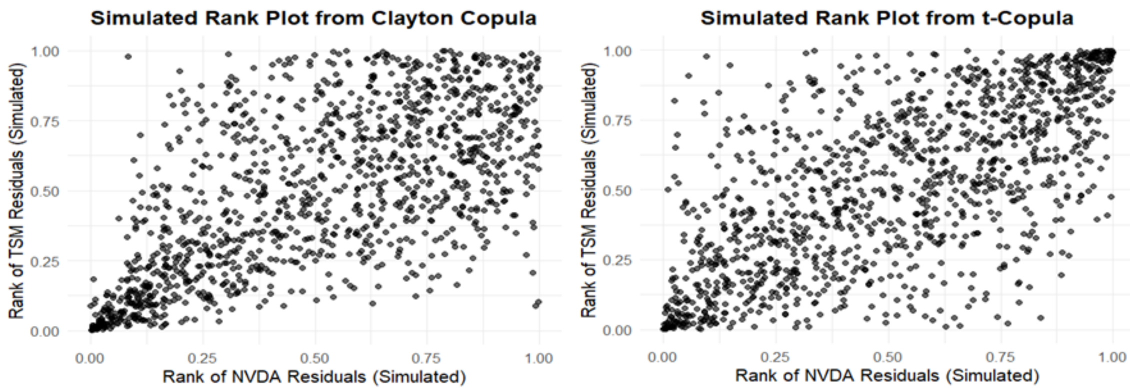

Figure 12: Rank Plots of simulated Clayton (left) and Student t (right) copula models with 1000 random data points.

The rank plots of our observed data and simulated t-copula plot look similar to each other, so that t-copula is the best bivariate copula model that measures the dependence of the residuals of NVDA and TSM stock returns among the five bivariate copula models listed in Table 1. This result is matched with other researches. For instance, in "The t Copula and Related Copulas" written by Demarta and McNeil in 2005, conclude that "The multivariate t-distribution is a natural choice for modeling asset returns due to its ability to account for heavy tails and tail dependence."

The tail dependence coefficient of the t-copula is defined as:

$$\lambda = 2t_{v+1}\left(-\sqrt{\frac{v+1}{1-\rho}}\right),$$

where v is the degrees of freedom, ρ is the correlation parameter which is our θ, and t_{v+1} is the CDF of the t-distribution with $v + 1$ degrees of freedom (Demarta & McNeil, 2005). Given the parameters $\theta = 0.67$ and $v = 6.33$, the tail dependence coefficient for t-copula $\lambda = 0.002$ (Appendix 6.3).

5 Conclusion

This study evaluated the dependence structure between the residuals of NVDA and TSM returns using copula models, specifically the t-Copula model. Traditional correlations such as Pearson correlation, Spearman's rho, and Kendall's tau, provided initial insights into the linear and monotonic relationships between these stocks. However, these quantiles fall short of considering complex dependence structures present in financial data.

The t-copula was selected for its ability to address these limitations, offering a more detailed view of the dependence structure in this data set. Unlike traditional correlations, the t-Copula can model tail dependence, which is the likelihood of extreme co-movements in both stocks' rates of returns. From the results, we know that Nvidia and Taiwan Semiconductor Manufacturing Company (TSMC) stock returns have a strong tendency to move together. In addition, as the degrees of freedom (v) increase, the t-copula behaves increasingly like a Gaussian copula with lighter tails (Demarta and McNeil, 2005). Thus, $v = 6.33$ in our t-copula model suggests a potential of moderate heavy tails. However, the estimated tail dependence coefficient ($\lambda = 0.002$) suggests that extreme co-movements between the NVDA and TSM are rare, which is consistent with the value of previous empirical (upper and lower) tail dependence coefficient. This finding demonstrates that while t-copula can accommodate moderate tail dependencies, the actual tail dependence in the dataset is relatively low.

From a hedging perspective to balance risk and earnings, the weak tail dependence obtained suggests that NVDA and TSM are unlikely to experience simultaneous extreme losses or gains. This finding implies that holding NVDA and TSM in a portfolio can provide diversified benefits, particularly during market shocks, as extreme joint losses are rare. However, the moderate linear dependence $\theta = 0.67$ between the two stocks shows the existence of regular co-movements, which may limit the hedging benefit under normal market conditions. Other similar pairs of firms that rely on each other can be further discussed to determine how this type of stock pairs should be handled in different market conditions.

In summary, the t-Copula offers a more complete and accurate representation of the dependency structure between NVDA and TSM than traditional correlation-based methods. By capturing both regular and extreme dependencies, the t-Copula enhances our ability to model and manage the risks associated with these assets in portfolio applications.

References

McNeil, Alexander J.; Frey, Rüdiger; Embrechts, Paul (2005). *Quantitative Risk Management. Concepts, Techniques and Tools*. Princeton Series in Finance, Princeton, NJ: Princeton University Press.

Cherubini, U., Luciano, E., & Vecchiato, W. (2004). *Copula Methods in Finance*. John Wiley & Sons.

Demarta, S. and McNeil, A.J. (2005). "*The t Copula and Related Copulas*". International Statistical Review, 73, 111-129.https://doi.org/10.1111/j.1751-5823.2005.tb00254.x

De Matteis, R. (2001). *Fitting Copulas to Data* (Diploma thesis). Institute of Mathematics, University of Zurich. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=1c55760c6fda7782f7fc13f666d5e79343b41656

Du, J., Lai, K.K. Copula-based risk management models for multivariable RMB exchange rate in the process of RMB internationalization. J Syst Sci Complex 30, 660–679 (2017). https://doi.org/10.1007/s11424-017-5147-3

Embrechts, P., Sidney I. Resnick, and Gennady Samorodnitsky.1999. "*Extreme Value Theory as a Risk Management Tool*", North American Actuarial Journal, Volume 3, Number 2, April 1999.

Embrechts, P., McNeil, A., & Straumann, D. (2002). *Correlation and dependence in risk management: Properties and pitfalls. In Risk Management: Value at Risk and Beyond* (pp. 176-223). Cambridge University Press. https://doi.org/10.1017/CBO9780511615337.008

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*(4), 987–1007. https://doi.org/10.2307/1912773

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics.*, *20*(7), 873–889. https://doi.org/10.1002/jae.800

Kendall, M. (1938). "*A New Measure of Rank Correlation*". Biometrika. **30** (1–2): 81–89. doi:10.1093/biomet/30.1-2.81

Mai, J F.; Scherer, M. (August 2012). *Simulating Copulas: Series in Quantitative Finance: Volume 4, Chapter 4.* World Scientific. https://doi.org/10.1142/p842.

McNeil, Alexander J.; Frey, Rüdiger; Embrechts, Paul (2005). *Quantitative Risk Management. Concepts, Techniques and Tools*. Princeton Series in Finance, Princeton, NJ: Princeton University Press.

Mukherjee S, Lee Y, Kim J, Jang J, Park J. 2018. *Construction of bivariate asymmetric copulas*. CSAM 2018; 25: 217-234.

Myers, Jerome L.; Well, Arnold D. (2003). *Research Design and Statistical Analysis* (2nd ed.). Lawrence Erlbaum. pp. 508. ISBN 978-0-8058-4037-7.

Nelsen, R.B. (2001) [1994], "Kendall tau metric", Encyclopedia of Mathematics, EMS Press.

Nelson, R.B. (2006). *An Introduction to Copulas*. New York, NY: Springer.

Patton, A.J. (2006). "*Modelling Asymmetric Exchange Rate dependence*". *International Economic Review*, 47: 527-556. https://doi.org/10.1111/j.1468-2354.2006.00387.x

Peng, Y., Ng, W.L. Analysing financial contagion and asymmetric market dependence with volatility indices via copulas. Ann Finance 8, 49–74 (2012). https://doi.org/10.1007/s10436-011-0181-y.

Ruppert, D. (2011). *Statistics and Data Analysis for Financial Engineering*. Chapter 8. Springer-Verlag. Retrieved from: https://faculty.washington.edu/ezivot/econ589/ch8-copulas.pdf

Sklar, A. (1959), "*Fonctions de répartition à n dimensions et leurs marges*", Publ. Inst. Statist. Univ. Paris, 8: 229–231.

Spearman, C. (January 1904). "*The Proof and Measurement of Association between Two Things*" (PDF). *The American Journal of Psychology*. **15** (1): 72–101. doi:10.2307/1412159. JSTOR 1412159.

Tim Bollerslev. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics. Volume 31, Issue 3, Pages 307-327. ISSN 0304-4076. https://doi.org/10.1016/0304-4076(86)90063-1.

Yahoo Finance. (n.d.). *Yahoo Finance Canada*. Retrieved September 24, 2024, from https://ca.finance.yahoo.com/.

Yoon, J. (2023, August 28). *Nvidia's critical weak point: Taiwan.* Financial Times. https://www.ft.com/content/bec85749-9354-4470-ac88-9323541c7bce.

Weisstein, Eric W. 2024. "*Statistical Correlation*." From *MathWorld*--A Wolfram Web Resource. https://mathworld.wolfram.com/StatisticalCorrelation.html.

Appendix A: Figures and Tables

Figure 6.

Histograms of NVDA and TSM stock returns.

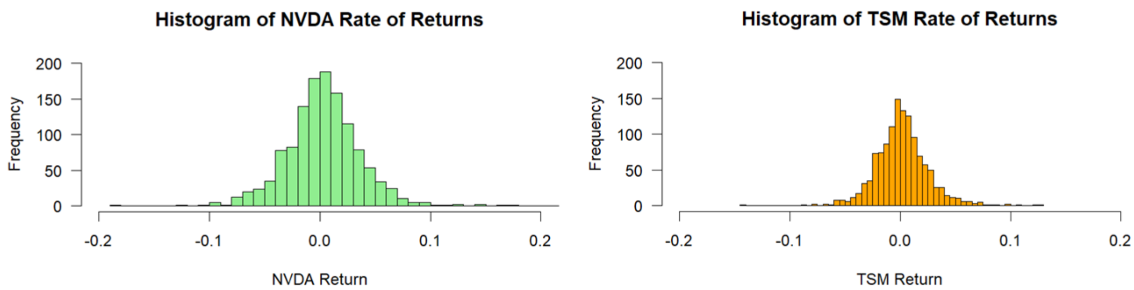

Figure 7.

Histograms of NVDA and TSM stock log-returns.

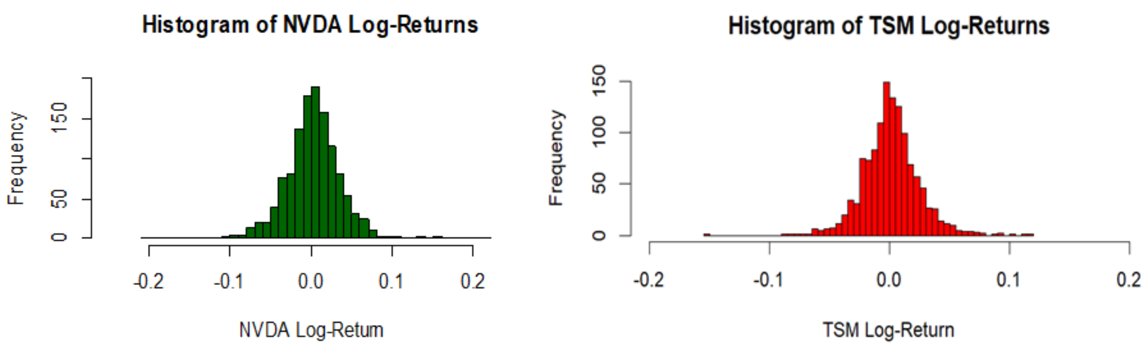

Table 1.

Few bivariate copulas.

Name	Distribution function	Parameter	Independence
Gaussian	$C(u,v) = \Phi_\theta(\Phi^{-1}(u), \Phi^{-1}(v))$	$\theta \in [-1,1]$	$\theta = 0$
t-copula	$C(u,v) = t_{\theta,v}(t_v^{-1}(u), t_v^{-1}(v))$	$\theta \in [-1,1]$, $v > 0$	$\theta = 0$
Clayton	$C(u,v) = [\max(u^{-\theta} + v^{-\theta} - 1, 0)]^{-\frac{1}{\theta}}$	$\theta \in (0, \infty)$	$\theta = 0$
Gumbel	$C(u,v) = \exp\{-\left((-\ln u)^\theta + (-\ln v)^\theta\right)^{\frac{1}{\theta}}\}$	$\theta \in [1, \infty)$	$\theta = 1$
Frank	$C(u,v) = -\frac{1}{\theta}\ln(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1})$	$\theta \in R, \theta \neq 0$	$\theta \to 0$

Table 2.

Statistical summary of raw data

Summary of Original Data (stock prices)		
Date	NVDA	TSM
Length: 1258	Min : 4.29	Min : 43.89
Class: character	1st Quartile : 13.13	1st Quartile : 80.26
Mode: character	Median : 19.30	Median : 96.05
	Mean : 31.26	Mean : 98.95
	3rd Quartile : 41.02	3rd Quartile : 118.24
	Maximum : 135.58	Maximum :191.05

Table 5.

Statistical Summary of Standardized Residuals of NVDA and TSM

Standardized Residuals						
	Minimum	1st Quartile	Median	Mean	3rd Quartile	Max
NVDA	-3.645	-0.577	-0.011	-0.003	0.601	9.571
TSM	-3.681	-0.583	-0.02	0.02	0.567	5.627

Appendix B: R Code

```
install.packages("readxl")  # The readxl package is widely used for reading Excel files
(.xls and .xlsx) into R
library(readxl)                        # Load the package
Data <-  read_excel("Data Analysis.xlsx")  # Import the dataset
str(Data)                              # See the structure of Data
summary(Data)                          # Summary statistics

# Calculate stock returns
library(dplyr)                 # Load the dplyr package to use its lag() function
Data <- Data %>% arrange(Time)   # Make sure the Data is sorted in chronological order.

# The lag() function in R is used to shift a vector or time series
Data$NVDAreturn <- (Data$NVDA - lag(Data$NVDA)) / lag(Data$NVDA)
Data$TSMreturn <- (Data$TSM - lag(Data$TSM)) / lag(Data$TSM)

# Log returns
Data <- Data %>% mutate(NVDAreturn_log = log(NVDA / lag(NVDA)))
Data <- Data %>% mutate(TSMreturn_log = log(TSM / lag(TSM)))

#Remove the NA value generated by shifting the rows
Data <- Data %>% filter(!is.na(NVDAreturn), !is.na(TSMreturn))

# Summary of the returns
summary(Data$NVDAreturn)
summary(Data$TSMreturn)

# Standard deviation of returns
sd(Data$NVDAreturn)
sd(Data$TSMreturn)

# Calculate correlations
cor(Data$NVDA,Data$TSM)                              # correlation of the price
cor(Data$NVDAreturn,Data$TSMreturn)                  # correlation of the return
cor(Data$NVDAreturn, Data$TSMreturn, method = "spearman")   # spearman's rho
cor(Data$NVDAreturn, Data$TSMreturn, method = "kendall")    # Kendall's tau

### Visualize the data
# Stock price by time
plot(Data$Time, Data$NVDA,
     xlab = "Time",
     ylab = "NVDA",
     main = "NVDA Prices")
plot(Data$Time, Data$TSM,
     xlab = "Time",
```

```
    ylab = "TSM",
    main = "TSM Prices")

# Stock returns by time
plot(Data$Time, Data$NVDAreturn,
    col = "darkgreen",
    xlab = "Time",
    ylab = "NVDAreturn",
    main = "Rate of retuns for NVDA")
plot(Data$Time, Data$TSMreturn,
    col = "darkred",
    xlab = "Time",
    ylab = "TSMreturns",
    main = "Rate of returns for TSM")

# Histogram of Returns
hist(Data$NVDAreturn,
    breaks = 60,
    col = "lightgreen",
    border = "black",
    main = "Histogram of NVDA Returns",
    xlab = "NVDA Return",
    ylab = "Frequency",
    xlim = c(-0.2, 0.2),
    ylim = c(0, 200),
    las = 1)
hist(Data$TSMreturn,
    breaks = 60,
    col = "orange",
    border = "black",
    main = "Histogram of TSM Returns",
    xlab = "TSM Return",
    ylab = "Frequency",
    xlim = c(-0.2, 0.2),
    ylim = c(0, 200),
    las = 1)

# Histograms of Log return
hist(Data$NVDAreturn_log,
    breaks = 50,
    col = "darkgreen",
    border = "black",
    main = "Histogram of NVDA Log-Returns",
    xlab = "NVDA Log-Return",
    ylab = "Frequency",
    xlim = c(-0.2,0.2),
```

```
    ylim = c(0, 200))
hist(Data$TSMreturn_log,
    breaks = 50,
    col = "red",
    border = "black",
    main = "Histogram of TSM Log-Returns",
    xlab = "TSM Log-Return",
    ylab = "Frequency",
    xlim = c(-0.2,0.2),
    ylim = c(0, 150))

#ggplot for returns
library(ggplot2)
ggplot(Data, aes(x = NVDAreturn, y = TSMreturn)) +
  geom_point(color = "blue", alpha = 0.6) +
  labs(title = "Scatter Plot of NVDA vs TSM Returns",
       x = "NVDA Returns",
       y = "TSM Returns") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, size = 16, face = "bold"),
        axis.text = element_text(size = 12),
        axis.title = element_text(size = 14))

#ggplot for log_returns
ggplot(Data, aes(x = Data$NVDAreturn_log, y = Data$TSMreturn_log)) +
  geom_point(color = "black", alpha = 0.6) +
  labs(title = "Scatter Plot of Log(NVDA) vs Log(TSM) Returns",
       x = "NVDA Log_returns",
       y = "TSM Log_returns") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, size = 16, face = "bold"),
        axis.text = element_text(size = 12),
        axis.title = element_text(size = 14))

### Example of Asymmetric Copula (Frank & Gumbel Copula)
library(copula)
gumbel_cop <- gumbelCopula(param = 5, dim = 2)    # Strong upper tail dependence for
diagonal concentration
frank_cop <- frankCopula(param = 2, dim = 2)      # Moderate dependence to add slight
banding

 # Set probabilities for sampling from each copula
p_gumbel <- 0.85  # 80% from Gumbel to emphasize the diagonal
p_frank <- 0.15   # 20% from Frank to add banding

 # Generate samples based on the mixture of copulas
```

```
k <- 1257
u_mixture <- matrix(0, k, 2)
set.seed(1215)
for (i in 1:k) {
  if (runif(1) < p_gumbel) {
    u_mixture[i, ] <- rCopula(1, gumbel_cop)  # Sample from Gumbel copula
  } else {
    u_mixture[i, ] <- rCopula(1, frank_cop)   # Sample from Frank copula
  }
}

  # Plot the result to show the banded diagonal sequence
plot(u_mixture, pch = 21, bg = "white", col = "black",
     xlab = "u", ylab = "v", main = "Frank & Gumbel Copula", xlim = c(0, 1), ylim = c(0,
1))

  # Optionally, add a diagonal reference line to emphasize the banding
abline(a = 0, b = 1, col = "red", lty = 2, lwd = 2)

### Simulated samples for 6 Copula model
n_samples <- 1000
rho_t <- 0.6         # t-Copula correlation
df_t <- 6            # t-Copula degrees of freedom
rho_gaussian <- 0.7  # Gaussian correlation
theta_clayton <- 2   # Clayton copula parameter
theta_gumbel <- 2    # Gumbel copula parameter
theta_frank <- 5     # Frank copula parameter

 # Create Copula models
sample_gaussian <- normalCopula(rho_gaussian, dim = 2)
sample_t <- tCopula(param = rho_t, dim = 2, df = df_t)
sample_clayton <- claytonCopula(theta_clayton)
sample_gumbel <- gumbelCopula(theta_gumbel)
sample_frank <- frankCopula(theta_frank)

set.seed(1215)
sample_independent_data <- cbind(runif(n_samples), runif(n_samples))  # Independent
case
sample_gaussian_data <- rCopula(n_samples, sample_gaussian)
sample_t_copula_data <- rCopula(n_samples, sample_t)
sample_clayton_data <- rCopula(n_samples, sample_clayton)
sample_gumbel_data <- rCopula(n_samples, sample_gumbel)
sample_frank_data <- rCopula(n_samples, sample_frank)

 # Combine all datasets and labels
```

```
sample_datasets <- list(sample_independent_data, sample_gaussian_data,
sample_t_copula_data, sample_clayton_data, sample_gumbel_data, sample_frank_data)
sample_titles <- c("Independent Case", "Gaussian Copula", "t-Copula", "Clayton
Copula", "Gumbel Copula", "Frank Copula")

 # Plot scatter plots
par(mfrow = c(2, 3))  # Arrange plots in 2 rows and 3 columns
for (i in 1:length(sample_datasets)) {
  plot(sample_datasets[[i]][, 1], sample_datasets[[i]][, 2],
       main = sample_titles[i],
       xlab = "u", ylab = "v",
       pch = 1 , col = "black", xlim = c(0, 1), ylim = c(0, 1))
}

#Fit with normal distribution
library(MASS)
fit_NVDA <- fitdistr(Data$NVDAreturn, "normal")
fit_TSM <- fitdistr(Data$TSMreturn, "normal")

# Transform the returns to uniform [0,1] using the CDFs of the fitted marginals
Uni_NVDA <- pnorm(Data$NVDAreturn, mean = fit_NVDA$estimate[1], sd =
fit_NVDA$estimate[2])
Uni_TSM <- pnorm(Data$TSMreturn, mean = fit_TSM$estimate[1], sd =
fit_TSM$estimate[2])

#Rank Plots
Data$Rank_NVDA <- rank(Data$NVDAreturn)  #Rank of returns
Data$Rank_TSM <- rank(Data$TSMreturn)
library(ggplot2)
ggplot(Data, aes(x = Rank_NVDA, y = Rank_TSM)) +
  geom_point(color = "blue", alpha = 0.6) +
  labs(title = "Rank Plot of Nvidia Returns vs TSM Returns",
       x = "Rank of Nvidia Returns",
       y = "Rank of TSM Returns") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, size = 16, face = "bold"),
        axis.text = element_text(size = 12),
        axis.title = element_text(size = 14))

# Tail Dependence
upper_quantile = 0.95
lower_quantile = 0.05
NVDA_Uthreshold <- quantile(Data$NVDAreturn, upper_quantile)
TSM_Uthreshold <- quantile(Data$TSMreturn, upper_quantile)
NVDA_Lthreshold <- quantile(Data$NVDAreturn, lower_quantile)
TSM_Lthreshold <- quantile(Data$TSMreturn, lower_quantile)
```

```
# Count how often both NVDA and TSM returns are greater than their 95th percentiles
upper_tail_events <- sum(Data$NVDAreturn > NVDA_Uthreshold & Data$TSMreturn >
TSM_Uthreshold)
# Count how often both NVDA and TSM returns are less than their 5th percentiles
lower_tail_events <- sum(Data$NVDAreturn < NVDA_Lthreshold & Data$TSMreturn <
TSM_Lthreshold)
m <- nrow(Data)

# Non-parametric tail dependence coefficient
upper_tail_dependence <- upper_tail_events / m  # upper tail
lower_tail_dependence <- lower_tail_events / m  # lower tail

### Apply the GARCH Time Series Model
install.packages("rugarch")
library(rugarch)

# Specify GARCH(1,1) model for NVDA returns
garch_NVDA <- ugarchspec(variance.model = list(model = "sGARCH", garchOrder =
c(1, 1)),
                  mean.model = list(armaOrder = c(0, 0), include.mean = TRUE),
                  distribution.model = "norm")
# Fit GARCH(1,1) model to NVDA returns
fit_garch_NVDA <- ugarchfit(spec = garch_NVDA, data = Data$NVDAreturn)

# Specify GARCH(1,1) model for TSM returns
garch_TSM <- ugarchspec(variance.model = list(model = "sGARCH", garchOrder = c(1,
1)),
                 mean.model = list(armaOrder = c(0, 0), include.mean = TRUE),
                 distribution.model = "norm")
# Fit GARCH(1,1) model to TSM returns
fit_garch_TSM <- ugarchfit(spec = garch_TSM, data = Data$TSMreturn)

###Residuals
# Extract raw residuals from both GARTH models and save in a numeric vector
res_NVDA <- as.numeric(residuals(fit_garch_NVDA))
res_TSM <- as.numeric(residuals(fit_garch_TSM))

# Extract standardized residuals (residuals divided by the estimated volatility)
stded_res_NVDA <- as.numeric(residuals(fit_garch_NVDA, standardize = TRUE))
stded_res_TSM <- as.numeric(residuals(fit_garch_TSM, standardize = TRUE))

# ACF plot to check for autocorrelation in standardized residuals
par(mfrow = c(1, 1))
acf(stded_res_NVDA, main = "ACF of Standardized Residuals (NVDA)")
acf(stded_res_TSM, main = "ACF of Standardized Residuals (TSM)")
```

```
# Ljung-Box test to statistically check for autocorrelation
Box.test(stded_res_NVDA, lag = 30, type = "Ljung-Box")
Box.test(stded_res_TSM, lag = 30, type = "Ljung-Box")

 # Print Summaries (Optional)
print(fit_garch_NVDA)
print(fit_garch_TSM)

### Tail Dependence for residuals
# Transform residuals to uniform [0,1] using empirical CDF
Rank_NVDAres <- rank(stded_res_NVDA)  #Rank of standardized residuals
Rank_TSMres <- rank(stded_res_TSM)
u_NVDAres <- Rank_NVDAres / (length(stded_res_NVDA) + 1)  #to avoid exact 0 or 1
u_TSMres <- Rank_TSMres / (length(stded_res_TSM) + 1)

# Set the quantile thresholds for upper and lower tails
U_quantile = 0.9
L_quantile = 0.1
UT_dep <- mean(u_NVDAres > U_quantile & u_TSMres > U_quantile)  #46 obs
LT_dep <- mean(u_NVDAres < L_quantile & u_TSMres < L_quantile)  #64 obs
print(UT_dep)                    # Upper tail dependence coefficient
print(LT_dep)                    # Lower tail dependence coefficient

## Examine significant of tail dependence
set.seed(1215)
n_permutations <- 10000
UT_perm <- numeric(n_permutations)
LT_perm <- numeric(n_permutations)
for (i in 1:n_permutations) {
  permuted_NVDA <- sample(u_NVDAres)
  UT_perm[i] <- mean(permuted_NVDA > U_quantile & u_TSMres > U_quantile)
  LT_perm[i] <- mean(permuted_NVDA < L_quantile & u_TSMres < L_quantile)
}

# p-values
U_p_value <- mean(UT_perm >= UT_dep)
L_p_value <- mean(LT_perm >= LT_dep)
print(U_p_value)  # p-value for upper tail dependence: p < 0.0001
print(L_p_value)  # p-value for lower tail dependence: p < 0.0001

# Standardized residuals for NVDA and TSM returns over time
par(mfrow = c(1, 1))
plot(Data$Time, stded_res_NVDA, type = "l", col = "green",
     main = "Raw Residuals of NVDA Returns Over Time", xlab = "Time", ylab =
"Residuals")
```

```
plot(Data$Time, stded_res_TSM, type = "l", col = "red",
    main = "Raw Residuals of TSM Returns Over Time", xlab = "Time", ylab =
"Residuals")

# Histogram of standardized residuals for NVDA and TSM
hist(stded_res_NVDA, breaks = 50, col = "lightgreen",
    main = "Histogram of Standardized Residuals (NVDA)", xlab = "Standardized
Residuals")
hist(stded_res_TSM, breaks = 50, col = "lightyellow",
    main = "Histogram of Standardized Residuals (TSM)", xlab = "Standardized
Residuals")

# Q-Q Plots
qqnorm(stded_res_NVDA, main = "Q-Q Plot of Standardized Residuals (NVDA)")
qqline(stded_res_NVDA, col = "red")
qqnorm(stded_res_TSM, main = "Q-Q Plot of Standardized Residuals (TSM)")
qqline(stded_res_TSM, col = "red")

# Rank plot for standardized residuals
library(ggplot2)
resrank_data <- data.frame(Rank_NVDAres = Rank_NVDAres, Rank_TSMres =
Rank_TSMres)
ggplot(resrank_data, aes(u_NVDAres, y = u_TSMres)) +
  geom_point(alpha = 0.6) +
  labs(title = "Rank Plot of Residuals",
       x = "Rank of NVDA Residuals",
       y = "Rank of TSM Residuals") +
  theme_minimal(base_size = 14) +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))

### Choose Copula Models
library(copula)
# Transform residuals to uniform [0,1] using empirical CDF
data <- data.frame(u_NVDAres, u_TSMres)
t_cop <- tCopula(dim = 2, dispstr = "un")  # unknown parameter
clayton_cop <- claytonCopula(dim = 2)
fit_t <- fitCopula(t_cop, data, method = "ml")
fit_clayton <- fitCopula(clayton_cop, data, method = "ml")

### Goodness of fit test
gof_t <- gofCopula(fit_t@copula, data, method = "SnC")
gof_clayton <- gofCopula(fit_clayton@copula, data, method = "SnC")
print(gof_t)
print(gof_clayton)

## Generate simulated plots from chosen copula model
```

```
n_points <- length(u_NVDAres)
t_copula_sim <- rCopula(n_points, fit_t@copula)
clayton_copula_sim <- rCopula(n_points, fit_clayton@copula)
t_copula_data <- data.frame(u_NVDAres = t_copula_sim[,1], u_TSMres =
t_copula_sim[,2])
ggplot(t_copula_data, aes(x = u_NVDAres, y = u_TSMres)) +
  geom_point(alpha = 0.6) +
  labs(title = "Simulated Rank Plot from t-Copula",
       x = "Rank of NVDA Residuals (Simulated)", y = "Rank of TSM Residuals
(Simulated)") +
  theme_minimal(base_size = 12) +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))
clayton_copula_data <- data.frame(u_NVDAres = clayton_copula_sim[,1], u_TSMres =
clayton_copula_sim[,2])
ggplot(clayton_copula_data, aes(x = u_NVDAres, y = u_TSMres)) +
  geom_point(alpha = 0.6) +
  labs(title = "Simulated Rank Plot from Clayton Copula",
       x = "Rank of NVDA Residuals (Simulated)", y = "Rank of TSM Residuals
(Simulated)") +
  theme_minimal(base_size = 12) +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))



# Tail dependence coefficient of t-copula
rho = 0.67
df_tcop = 6.33
lumda_t <- 2 * pt(-sqrt((df_tcop + 1) / (1 - rho)), df = df_tcop + 1)
print(lumda_t)
```