London School of Economics and Political Science

ST429 - Statistical Methods for Risk Management

Statistical Methods for Green Finance Analysis

Group 6

Candidate Number: 36718, 29393, 40168, 21851

January 6, 2022

Abstract

Due to the Paris Agreement and the United Nations Climate Change Conference (COP26), many investors are seeking signs of how a rapid transition to a zero-carbon economy will affect their investment orientation. Hence, we analysed some related stocks using different statistical methods. In stock price analysis, volatility clustering of each stock that may be caused by the whole economic market, the industry it belongs to, or its unique performance could be found out. For four relevant pairs of stocks, the dependence structures of the different risks are modelled by various copulas and the best fit copula is determined. In the principal component analysis of stock returns, the first principal component could be identified as the market factor.

1 Introduction

Climate change has been a trending topic over the past few years, arising the attention of countries around the world. In December 2015, a total of 195 countries signed the Paris Agreement during the COP21 in France, aiming to reduce the risk of the increasingly serious issue. The significance of such agreement can not only be implemented by individuals themselves, such as by reducing the usage of plastic items and adopting rubbish classification, but also by the society as a whole. In commercial aspects, for example, the climate-related markets may seem to be intensively influenced by the rising attention of environmental protection. Some businesses such as renewable energy companies may benefit from this green regulation through means of subsidies or tax deductions from the government, which would hopefully promote their development, whereas some high emission firms may be thus negatively affected.

The objective of this report is to analyse ten different stocks selected from several industries, which are related to the two sides of the green policy, either positively or negatively. Together with the S&P500 index, we evaluate their performance through various statistical methodologies. In particular, the background information of our chosen stocks is introduced, including justifications of why they are classified as good stocks or bad stocks in Section 2. Section 3 discusses the log returns of each stock and the obvious volatility that we find during our analysis. Section 4 is devoted to a discussion of risk measures assuming empirical distribution and Pareto distribution, with the comparison of various capital requirements resulted. Next, Section 5 presents the dependence structure between four stock pairs using various copulas and find the best fit of each pair among those copulas. In Section 6, the results from the previous part are utilised to find an estimated joint distribution of a chosen stock pair, and Gumbel copula is used instead to compare the joint model. Finally, in Section 7, a principal component analysis is constructed to determine the copula that best fits the data.

2 Stock Background

The stocks we will discuss in this report are shown in the following Table 1, including their company names, industries they are related to, and volatility of the stock as compared to the whole market (denoted as the Beta index in finance). The stocks that are positively connected with the green policy are marked in green, while those negative ones are in grey. In this report, we focus on several sectors that are considered to be closely connected to environmental behaviour and treat companies that would benefit from the COP26 Climate

Talk as winners, whereas those that would suffer as losers. The corresponding stocks are thereby classified as good and bad stocks.

Stock Symbol	Company Name	Industry	Beta
AWK	American Water Works Company	Water	0.28
FTEK	Fuel Tech	Pollution Control	5.71
GE	General Electric Company	Wind Power	1.04
TSLA	Tesla	Transportation	2.04
WTRG	Essential Utilities	Water	0.57
AAL	American Airlines Group	Airlines	1.67
ED	Consolidated Edison	Electric	0.20
EGLE	Eagle Bulk Shipping	Marine Shipping	1.30
F	Ford Motor Company	Transportation	1.09
XEL	Xcel Energy	Electric	0.37

Table 1: Information of stocks selected

According to Bloomberg Green, the COP26 Climate Talk favouring renewable energy would essentially benefit the green industry while adding pressure to the traditional one. Take the water industry as an example, Essential Utilities (WTRG) acts as a primary water trader in the States; while as the largest United States water utility company, American Water (AWK) constantly supplies potable water to around 14 million people all over the country. When it comes to the wind sector, General Electric Company (GE), as a representative of initiating sustainable development, takes advantage of the wind to supply the world's economy. Nevertheless, companies providing electric and natural gas services, such as Consolidated Edison (ED) and Xcel Energy (XEL), are considered to be victims of the act regarding the source of energy they adopt for their services.

In order to move towards a territory with ideally zero carbon emission, it is commonly believed that pollution control plays an important role. As one of the industry leaders, Fuel Tech (FTEK) benefits from its strength in technology and devotes consistently to mitigating greenhouse gases. This, on the other hand, is also the goal of Tesla (TSLA), an electric vehicle manufacturer dedicated to developing alternative means of transportation. In this case, it is expected that these two companies would hopefully benefit from the release of the new document.

Speaking of transportation, the traditional industry seems to suffer from the termination of the summit due to its unfriendliness for the environment. Ford Motor Company (F), for example, has focused on the production of gasoline cars since its foundation and may thereby be adversely affected by the effectiveness of the agreement, which aims to control carbon dioxide emission. The same problem would arise in the airline and shipping industry, and in this report, stocks from American Airlines Group (AAL) and Eagle Bulk Shipping (EGLE), representatives from the two sectors, are analysed. Specifically, American Airlines Group acts as one of the world's largest airlines and offers over six thousand flights to more than 300 places in the world, while Eagle Bulk Shipping uses its Supramax and Ultramax ships and develops its business in the world dry bulk shipping. It is thus worth considering whether the restriction of emitting greenhouse gases would essentially affect the performance of such enterprises.

3 Stock Prices Analysis

This section focuses on the analysis of the log returns for the stocks in our portfolio. In this section, Package {xts} is used to process data, and the log returns for the stocks are shown in Figure 1, with good stocks on the left and bad stocks on the right. Throughout our analysis, the stock prices are evaluated not only with respect to individual companies, but also to the relating industry sectors and the market as a whole, so that we would be able to observe relationships among these stocks.

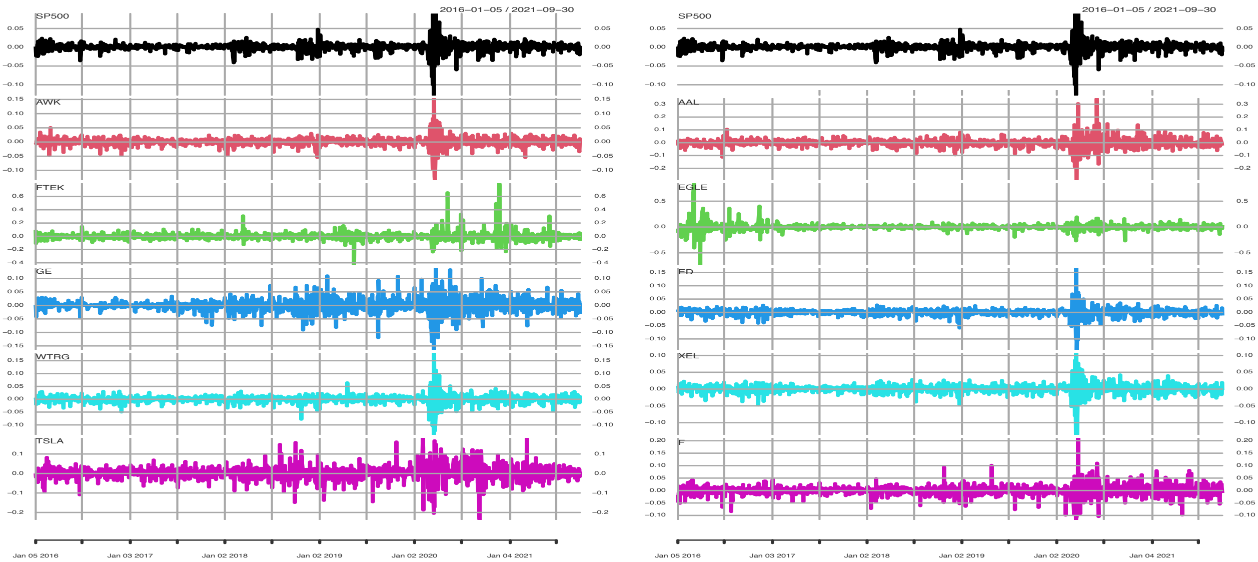


Figure 1: Log returns for ten stocks

Compared with the log returns of the S&P 500 index, the stock market is correlated to a degree due to some big events that fluctuate the whole economy. For instance, in December 2018, the launching of America's trade war with China brought the risk of uncertainty in future interest rates. As a result, the S&P 500 index fell by 11% in that month and continued to be unstable in a period of time. On 16 March 2020, nicknamed "Black Monday II", the S&P 500 index plunged 12% due to the globally spread of COVID-19. Fortunately, the market gradually returned to normal starting from the second quarter of 2020.

Aside from the whole market movements, the volatility of stocks from companies in similar sectors usually arise simultaneously. For instance, in the electricity generating sector, the stock prices of XEL and ED went down by 7% and 6% in the same month respectively for the reason that natural gas overtook coal in the electric power industry for the first time, provided by World Energy Outlook 2018. In addition, according to the Stated Policies Scenario (STEPS), renewable technologies will capture more than three quarters of the electricity market by 2040. In the transportation sector, the whole industry was facing a supply chain crisis due to the semiconductor chip shortage at the beginning of 2021. Thus, we can see that both TSLA and F experience some extent of volatility during that time.

Extreme return points are also visible in terms of individual stocks. For example, a ruling by the European Court of Justice was released on November 10, 2020, saying that "Italy 'systematically and persistently' breached EU rules on air pollution", and "entitled to levy fines against Italy." As one of the Fuel Tech, Inc.'s main business countries, the EU's move may trigger Italy's own attention to air pollution, causing the skyrocketed increase of FTEK, which showed the highest record of 0.797 log return in one day. What's more, the stock price of AAL experienced some fluctuations around 2016 and 2020. According to Fortune, in the first quarter of 2016, though the low oil prices provided airlines with more cash flow, they contributed to better allocation of seats and lower ticket prices for rival carriers of American Airlines, leading to a fall in sales. Also, the passenger revenue per available seat mile, as a result, fell, while the following rising oil prices decreased the unit revenue and led to the margin contraction, according to The Motley Fool.

However, by comparing historical data of the positive side stocks and the negative side stocks, there seems to be no obvious trend for the former to go upwards after the signing of the Paris Agreement. One of the most significant reasons may be that there has been no detailed policies set to implement the agreement. Hence, the level of engagement from different countries is hard to quantify. In view of the COP26 Summit in Glasgow, we expect the green stocks to go up since all countries are supposed to give plans on how to reduce the emission of greenhouse gases and the government will put more effort to support those green companies.

4 Loss and Risk Measures of the Portfolio

4.1 Loss of the Portfolio

Stock	**Price** ($S_{0,i}$)	**Price** ($S_{t,i}$)	**Weight** ($w_{t,i}$)
AWK	54.09	168.434	12.84%
FTEK	1.8	1.74	0.13%
GE	213.06	103.03	7.86%
TSLA	44.68	775.48	59.13%
WTRG	25.67	45.82	3.49%
AAL	39.10	20.52	1.57%
ED	51.61	71.88	5.48%
EGLE	61.86	47.98	3.66%
F	10.58	14.09	1.07%
XEL	30.09	62.5	4.77%

Table 2: Adjusted Close Price and Weights of Stocks

Now that we have developed methods to analyse adjusted close prices and log returns of each stock in our portfolio, in this section, we are ready to evaluate the random losses. Consider a share of 1000 for each stock in our portfolio, we obtain an initial investment of

$$V_0 = \sum_{i=1}^{10} 1000 * S_{0,i}$$

where $S_{0,i}$ represents the adjusted close stock price for stock $i = 1, \ldots, 10$ at the time 0, and V_0 is the value of the portfolio at that time.

Table 2 shows the adjusted stock price in our portfolio at time 0 (January 4th 2016) and time t (current time, September 30th 2021). From the table, we can get the initial investment of $903664.99. Similarly, a total value of $1311471.67 is obtained at the end of the period. Furthermore, we can compute the weights of each individual stock in the portfolio by current prices, i.e.

$$w_{t,i} = 1000 * \frac{S_{t,i}}{V_t}$$

where $S_{t,i}$ is the adjusted close price at current time and $V_t = \sum_{i=1}^{10} S_{t,i}$ is the total value of investment now.

The random loss associated to the portfolio is then constructed. By theory, the random loss at time t+1 (next day) is defined by the negative of change in value of the portfolio from time t+1 to t. In our case, we build the random loss for next day with individual weights and log returns, and we get

$$L_{t+1} = -V_t \sum_{i=1}^{10} w_i(e^{X_{t+1,i}} - 1))$$

where $X_{t+1,i} = \log S_{t+1,i} - \log S_{t,i}$ is the difference of log-returns at t+1 and t, known as risk factor change.

4.2 Distribution of Positive Losses

In this report, we are only interested in evaluating the losses of our portfolio, so the positive random losses of our stocks are filtered, and Table 3 shows the summary of how the losses behave. The frequency of values is also plotted in a histogram (Figure 2). According to the results, it is suggested that though the losses have a relatively large spread, from the lowest $61.3 to maximum of $203404.4, the amount of loss is concentrated around $10000 to $20000, with less frequency as the amount increases.

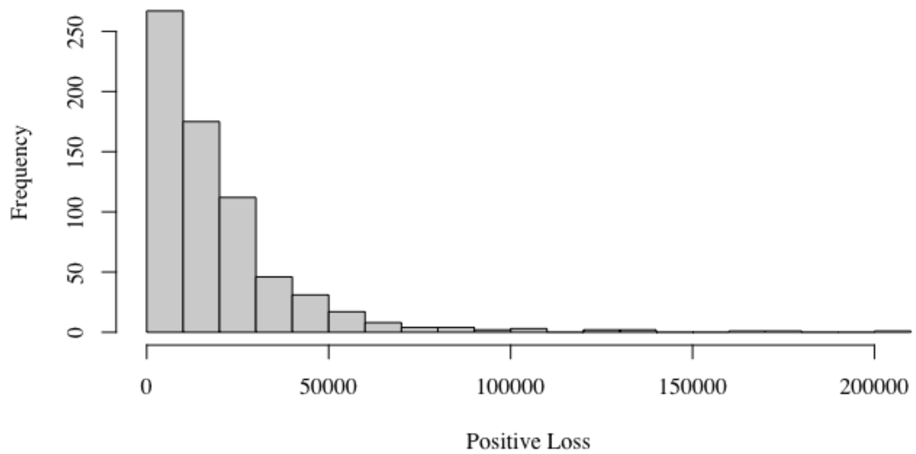


Figure 2: Histogram of Positive Losses

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
61.3	5722.7	13681.8	19502.0	25018.7	203404.4

Table 3: Summary of Positive Loss

From the figure, we can also see that there is hardly any frequency of positive losses as the amount getting extreme. Thus, we can find that there is a right-skewed pattern shown in the shape of positive losses. So considering the tail on the right of the distribution, a Pareto model is fitted to better evaluate the tail behaviour. According to the definition, a Pareto distribution $Pa(\alpha, \kappa)$ has cumulative distribution function of

$$F(x) = 1 - (\frac{\kappa}{\kappa + x})^{\alpha}$$

for $\alpha > 0$ as the shape parameter, $\kappa > 0$ as the scale parameter and $x \geq 0$.

The R function fitdist() in the {fitdistrplus} package is used to fit the distribution, and after calculation, in our problem, we get the values $\alpha = 11.22$ and $\kappa = 198121.59$ for the portfolio. Then, two risk measurement approaches, Value-at-Risk (VaR) and Expected Shortfall (ES), are taken to evaluate the capital requirement for this portfolio.

4.3 Capital Requirements with VaR and ES

Now that we have fitted the Pareto distribution on our data, in this part, assuming the empirical or Pareto distribution, we evaluate the Value-at-Risk (VaR) and Expected Shortfall (ES) at various confidence levels, and the results can be shown in Figure 3 below.

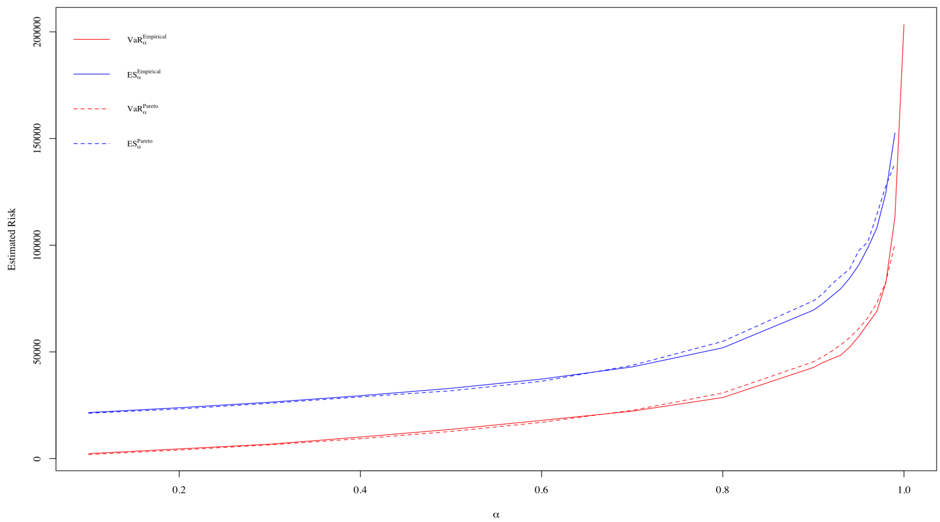


Figure 3: VaR and ES at different levels

From the graph, it is shown that both empirical and Pareto distributions produce a similar amount of VaR and ES, especially for small confidence levels, when the solid and dashed lines coincide. As the confidence level increases, however, starting from approximately 70%, the Pareto distribution results in a slightly greater loss than assuming the empirical distribution for our observations. Nevertheless, when the confidence level gets large enough to 1, we observe that the empirical distribution would essentially lead to a higher risk of loss. Unsurprisingly, the ES calculated turns out to be greater than VaR for both models, which is consistent with the definition. Since ES is the expected value of the losses greater than the VaR calculated, it looks further into tails and would always be larger than VaR. We further evaluate the capital requirements with VaR and ES at a level of 90%, 95% and 99% specifically, and the numbers are summarised in Table 4 and 5 respectively.

According to the tables, at a confidence level of 90% and 95%, the risk of loss associated with our portfolio is estimated to be higher when we assume a Pareto distribution than the empirical one; while the number is smaller when we consider a confidence interval of 99%, resulting in less capital requirements needed to compensate for the loss. Also, as we expected, the ES measure produce a greater amount of value than the VaR method at all levels.

	90%	95%	99%
Empirical	42682	57212	113050
Pareto	45356	60859	100620

Table 4: Capital Requirement with VaR

	90%	95%	99%
Empirical	69557	90680	152619
Pareto	73951	97435	138409

Table 5: Capital Requirement with ES

5 Dependence Structure

In this section, package {QRM} is used to construct pseudo copula data and package {copula} is used to fit copulas. For selecting copulas, function BIC() is applied.

5.1 Rank Correlation

In this part, we created four sample pairs of our stocks based on their sectors and computed the corresponding rank correlations Kendall's Tau and Spearman's Rho. The results are summarised in Table 6 below. From the table, it is obvious that there exist some positive correlations between all pairs of stocks. In particular, the first pair (XEL-ED) and the third pair (AWK-WTRG) both show a relatively high rank correlation in accordance with our suggestion in Section 3 that stocks in the same sector tend to experience similar changes and move together. However, we can see that there is some inconsistency in the transportation sector, where TSLA and F show a low correlation. This may be caused by the fact that TSLA focuses on electric car services while F concentrates on traditional fuel vehicles, as mentioned in the introduction of stock background (Section 2). For the last pair (AAL-EGLE), despite the fact that they belongs to different industries, we group them into pairs since they have a certain degree of service intersection, as airlines can be treated as one of the shipping methods. In this case, though the correlation is relatively low, one cannot consider them as independent.

	XEL-ED	TSLA-F	AWK-WTRG	AAL-EGLE
Kendall's Tau	0.634472	0.143171	0.607525	0.183129
Spearman's Rho	0.809871	0.207142	0.785224	0.265601

Table 6: Rank Correlation

5.2 Copula Families

A copula is a probability integral transform model widely used in finance for observing the dependence structure directly without the computation of joint and marginal distributions and is especially helpful when dealing with high dimensional data. As shown in table 7, in this section, we consider Gaussian copula, t copula, Gumbel copula, Clayton copula and Frank copula. Specifically, Gaussian copula is a classic model dealing with linear correlation but may break down in tail dependence. t copula, on the other hand, has a heavier tail than the Gaussian copula, and as the degree of freedom increases, it converges to the Gaussian copula. The rest three copulas are all Archimedean copulas with a closed form of the generator function. Frank copula is the only symmetric copula, while Gumbel copula has greater upper tail dependence and Clayton copula has greater lower tail dependence.

Copula	**Formula**
Gaussian	$C_P^{Ga}(u_1, u_2) = \Phi_P(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$
t	$C_{\nu,P}^{t}(u_1, u_2) = t_{\nu,P}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2))$
Gumbel	$C_\theta^{Gu}(u_1, u_2) = exp\{-((-lnu_1)^\theta + (-lnu_2)^\theta)^{\frac{1}{\theta}}\}$
Clayton	$C_\theta^{CL}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}$
Frank	$C_\theta^{Fr}(u_1, u_2) = -\frac{1}{\theta} ln(1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1})$

Table 7: Copula Families

5.3 Methods to Estimate Copulas

In this report, the Pseudo maximum likelihood estimation (MLE) method is applied, and the algorithm to compute pseudo MLE is shown below. Since the MLE function in R is based on an iteration of parameters to convergence, setting start points could help to reduce the number of iterations. In particular, the inverse of Spearman's Rho is used as a start point for Gaussian copulas, while the inverse of Kendall' Tau is used as a start point for Archimedean copulas. Note for t copula, as the degree of freedom cannot be estimated, we do not give a start point.

Algorithm 1 Pseudo MLE method

Input: Observations $Y_1, Y_2, ..., Y_n$;

Output: Fitted parameter $\hat{\theta}$;

1: Fit data to empirical distribution$\hat{F}_{Y_j}(y) = \frac{1}{n+1}\Sigma_{i=1}^n \mathbb{1}_{y_{i,j} \leq y}$

2: Compute $lnL(\theta_c) = \Sigma_{i=1}^d ln(c_Y(\hat{F}_{Y_1}(y_{i,1}), ..., \hat{F}_{Y_d}(y_{i,d})))$

3: Find $\hat{\theta}_C$ that maximises $lnL(\theta_c)$

5.4 Selection Criterion

In order to compare the quality of fitted models, the Akaike information criterion (AIC) is introduced. The AIC value is computed by

$$\text{AIC} = 2k - 2\ln(\hat{L})$$

where k is the number of estimated parameters and $\hat{L}$ is the maximum likelihood function. It is known that AIC is an adjusted likelihood function that penalises the complexity of the model, so a model with a lower AIC score is preferred.

5.5 Fitting Copula

The four pairs of stocks introduced in Section 5.1 are fitted to Gaussian copula, t copula, Gumbel copula, Clayton copula and Frank copula separately. In each copula family, the pseudo MLE is applied to find the parameter of best fit, while among different families, AIC is used to compare the goodness of fit. Additionally, simulations are drawn from the selected copula (Figure 4) to test the model. The summary of fitted copulas is shown in Table 8 and Table 9. Note that Clayton copula is not applicable for TSLA-F and AAL-EGLE, as the parameter does not converge in log-likelihood maximisation.

XEL-ED	max. loglik	AIC	**TSLA-F**	max. loglik	AIC
Gaussian	809.093	-1616.19		37.9873	-73.9747
t	923.426	-1842.85		67.0934	-130.1867
Gumbel	847.516	-1693.03		56.5044	-111.0087
Clayton	697.125	-1392.25		NA	NA
Frank	789.111	-1576.22		32.8363	-63.6725
AWK-WTRG	max. loglik	AIC	**AAL-EGLE**	max. loglik	AIC
Gaussian	747.741	-1493.48		55.8106	-109.621
t	836.333	-1668.67		67.4306	-130.861
Gumbel	801.751	-1601.50		62.9164	-123.833
Clayton	587.846	-1173.69		NA	NA
Frank	710.959	-1419.92		53.0127	-104.025

Table 8: Fitted Copulas

	Best Copula	λ_l	λ_u
XEL-ED	t	0.606244	0.606244
TSLA-F	t	0.124134	0.124134
AWK-WTRG	t	0.564789	0.564789
AAL-EGLE	t	0.0650392	0.0650392

Table 9: Lambda for Best Fitted Copula

In XEL-ED, TSLA-F, AWK-WTRG and AAL-EGLE, t copula has Kendall's Tau of 0.6338, 0.1447, 0.6069 and 0.1831 respectively, which is accurate to at least two decimal places compared with sample correlation (Table 6). However, the discrepancies may be more significant for Spearman's Rho due to the reason that Spearman's Rho is more error sensitive. As shown in Table 8, surprisingly, t copula has the lowest AIC among all four pairs. The log return graph (Figure 1) in Section 3 may explain this situation since almost all stocks experience their highest volatility during the first quarter of 2020 due to COVID-19. Hence, it can be deduced that all stocks have higher tail dependence than normal, so t copula is reasonable, as it has equal dependence in the upper tail and lower tail. If we fit AAL-EGLE with data before June 2019, we can find that Gaussian copula gives a lower AIC score. This indicates that the correlation of AAL-EGLE is no longer significant, excluding the COVID-19 period.

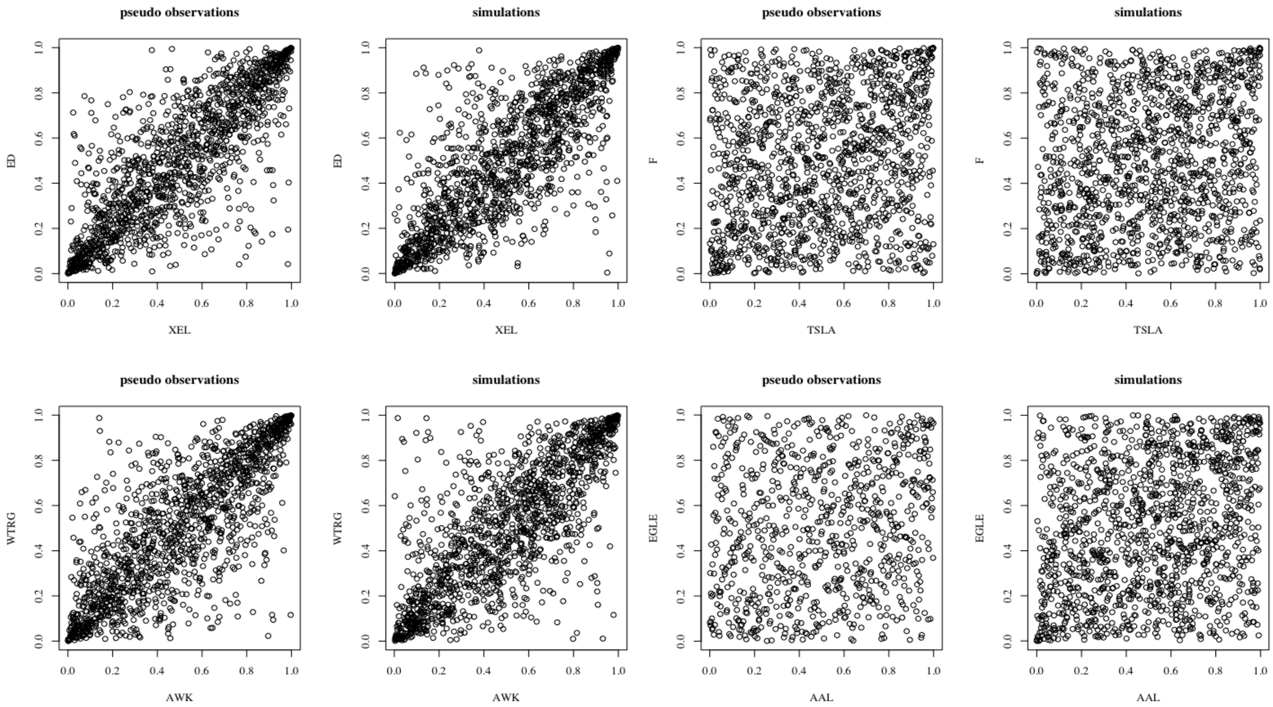


Figure 4: Simulations of Best Fitted Copula

6 Loss and Risk Measures of Bivariate Pair

In this section, R package {actuar} is used for taking Pareto distribution into account and package {fitdistrplus} is used to construct a fitted distribution for marginals. Function mvdc() from package {copula} is also used to derive the joint distribution.

6.1 Joint Model Estimation

In order to estimate the joint model, both the copula and marginal distributions are needed. From the previous section, we know that the best fit is the t copula, and thus only marginals need to be further considered. Typically, the first attempt is to use the empirical distribution to compute an estimate of marginals. However, the derived marginals are discrete and also the cumulative distribution function is always equal to one if the value is greater than the maximum point in the data. These findings are unrealistic to the real-world cases, so other methods should be considered.

Since the log-returns are leptokurtic, Pareto distribution, a heavy-tailed distribution that only takes positive losses into account, would possibly be a better choice. For a chosen stock pair XEL and ED, we use the function fitdist() to compute its fitted marginal distribution and combine it with the t copula. The function mvdc() is used to derive the bivariate joint model. We then draw random numbers to get simulated data and compare them with original data, as shown in Figure 5. From the graph, except for a few outliers, these two plots show a similar pattern, showing that we have a reasonably appropriate estimated bivariate model.

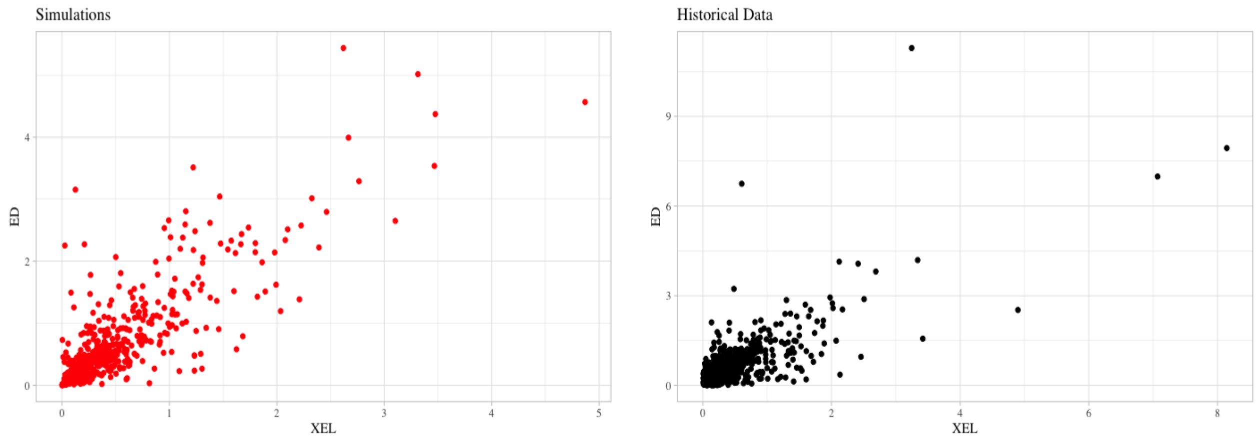


Figure 5: Bivariate Model Test

6.2 Expected Shortfall of the Aggregated Losses

In this part, Monte Carlo method is used to find the Expected Shortfall of the aggregated losses. The basic Monte Carlo integration uses random sampling of a function to numerically compute an estimate of its integral. This is based on the law of large numbers (LLN), which can be interpreted as $\frac{1}{n}\Sigma_{i=1}^{n}X_i \xrightarrow{P} \int xf(x)$, if $X_1, \ldots, X_n$ are samples from the distribution with probability density function f. Then for any function h, we apply the LLN to the sample H_i=$h(X_i)$, leading to

$$\overline{H}_n = \frac{1}{n}\Sigma_{i=1}^{n}h(X_i) \xrightarrow{P} E[h(X_1)] = \int h(x)f(x)dx.$$

By definition, $ES_\alpha(L)$=$E[L \mid L \geq VaR_\alpha(L)]$, which means that expected shortfall is the conditional expectation of loss given that the loss is beyond the Value-at-risk (VaR) level. Thus, let $h(X_1)$=$\{L \mid L \geq VaR_\alpha(L)\}$ and $f(x)$ be the probability density function of loss L, the Monte Carlo simulation could be applied.

Now, we compute 1000 pairs of losses randomly from our estimated joint model using function rMvdc() to get the respective losses L_1 and L_2 and sum these two losses in each pair to get the aggregated losses. Next, we look for the quantiles of the aggregated losses and thereby deriving the expected shortfall using the equation above.

Assuming the pair XEL and ED has a Gumbel copula with parameter $\theta \to \infty$, the limit of C_θ^{Gu} is the comonotonicity copula M. Then, we get $F_{L_1+L_2}^{\leftarrow}(\alpha) = F_1^{\leftarrow}(\alpha)+F_2^{\leftarrow}(\alpha)$, meaning that the VaR of the aggregated risk is the sum of the individual VaR, same for the Expected Shortfall as a result. Thus, 1000 samples of each loss are selected randomly by function rpareto() and the sum of their individual Expected Shortfall is computed to derive the Expected Shortfall of the aggregated losses.

The following table 10 shows the Expected Shortfall (ES) values with t copula, Gumbel copula and the original data respectively at different confidence levels.

α	0.2	0.4	0.6	0.8	0.9	0.95	0.99
ES with t Copula	1.4505	1.7804	2.2584	3.1245	4.0465	5.0447	7.7488
ES with Gumbel Copula	1.4646	1.8055	2.3346	3.2899	4.4797	5.8812	7.8052
ES of data	1.6152	1.9564	2.4421	3.3784	4.5490	6.1139	11.1613

Table 10: Expected Shortfall of the Aggregated Losses

A t copula is an elliptically symmetric distribution with heavy left tails. Gumbel copula, which is asymptotically independent, exhibits more weight on the positive tail than the negative. Thus, both of the two copulas allow the modelling for tail dependence between risks.

From the table 10, it is shown that the Expected Shortfall with t copula and Gumbel copula is relative to the same at low confidence levels. However, when confidence levels become larger, particular at 0.9 and 0.95, the results of the Gumbel copula have greater values compared to the t copula, which is caused by the fatter tails Gumbel has. Since the Expected Shortfall is more sensitive to the shape of the tail of the loss distribution, the usage of Pareto distribution as a marginal fit weaken the tail losses to some extent, which means the original data has fatter-tailed losses. As a result, it can be concluded that both of the estimated Expected Shortfall values are underestimated.

7 Principal Component Analysis

7.1 Principal Component Analysis for Log Return

From the previous analysis, it is expected that stocks would simultaneously suffer changes and opportunities under the same background. As a result, Principal Component Analysis is applied to reduce dimensions and deal with hidden information. Upon implementation, we extract the first principal component with the most critical elements and search for the dependence structure between the item and the second crutial component.

Algorithm 2 Principal Component Analysis

Input: High dimension multivariate observations $X_1, X_2, ..., X_n$;

Output: Low dimension transforms $Y_1, Y_2, ..., Y_d$;

1: Calculate covariance matrix $S_x = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(X_i - \bar{X}\right)'$

2: Do spectral decomposition $S_x = GLG'$, where L contains ordered eigenvalues λ_i and G contains corresponding eigenvectors

3: Convert principal components $Comp_i = g_i'(X - \bar{X})$

4: Show cumulative contribution of k-th principal components $\sum_{i=1}^{k}\lambda_i / \sum_{i=1}^{d}\lambda_i$

By adopting the prcomp() function in the basic {stats} package, we carry out the Principal Component Analysis. The results are shown in Table 11 below. According to the table, the first five principal components accumulate a total contribution of 93.8%, and the first principal component (known as Comp.1) explains 37.62% of the variation among all ten principal components. It is also worth mentioning that the coefficient of all related features to Comp.1 are positive, suggesting that the ten stocks serve a positive relationship as a whole.

Stock	AWK	FTEK	GE	WTRG	TSLA	AAL	EGLE	ED	XEL	F	Cumulative
Comp.1	0.0129	0.4080	0.1265	0.0371	0.1281	0.1835	0.8655	0.0106	0.0103	0.1285	37.62%
Comp.2	-0.0095	0.8959	0.0031	-0.0027	0.0533	0.0334	-0.4394	-0.0056	-0.0118	0.0135	65.98%
Comp.3	0.1241	-0.1711	0.3300	0.1489	0.5681	0.5763	-0.2310	0.0934	0.0209	0.3092	79.46%
Comp.4	-0.0452	-0.0009	-0.2177	-0.0568	0.8033	-0.5146	0.0522	-0.0688	-0.0956	-0.1709	87.87%
Comp.5	0.4643	0.0354	0.1546	0.4963	-0.0863	-0.3829	0.0094	0.4217	0.0164	0.0328	93.8%

Table 11: Contribution of first five Principal Components

7.2 Dependence Structure of Whole Period

From the results of the contribution of the first Principle Components, it is easy to observe that the stock FTEK attain the maximum in Comp.2 in terms of absolute values (0.8959). Hence, in this part, we concentrate on investigating the dependence structure between Comp.1 and the stock FTEK. Intuitively, we expect the copula to be independent, as the principal components are orthogonal. By performing the asymptotic independence test, we obtain that $T = 18.0725$ and $p-$value $= 0$, revealing the correlation between the two. Therefore, it seems to be arbitrary to claim the existence of an independent copula.

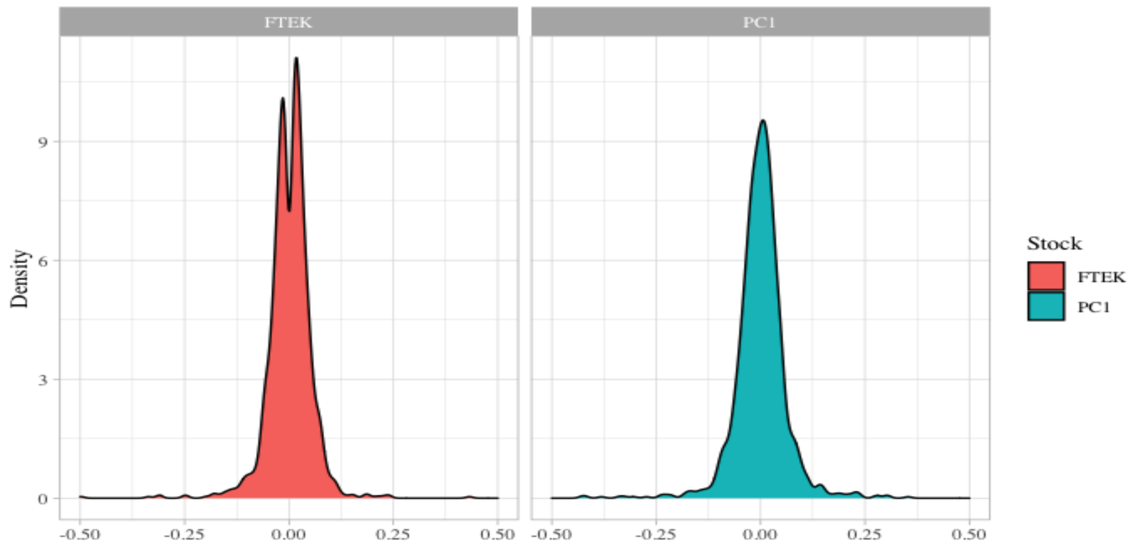


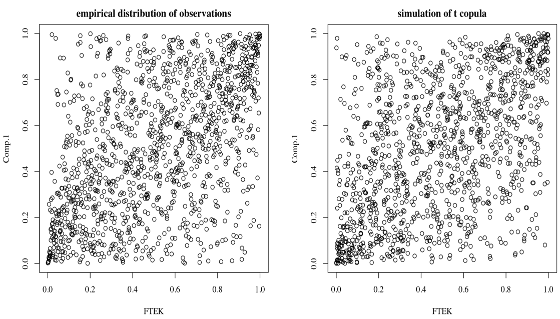
Figure 6: PDF of FTEK and Comp.1

The kernel density distribution of FTEK and Comp.1 is plotted in Figure 6. From the graph, both distributions are showing the pattern of centralisation and symmetry. However, the figure also shows the existence of a subtle difference. Specifically, the density of Comp.1 is more likely to be a normal distribution, while the density function of FTEK is partially depressed as approaching 0. The Pseudo MLE method is then applied to estimate the copula, and the AIC statistics are used to evaluate the quality, as discussed in Section 5. Finally, various copulas, Gaussian copula, t copula, Gumbel copula, Clayton copula and Frank copula, are fitted to determine the best dependence structure.

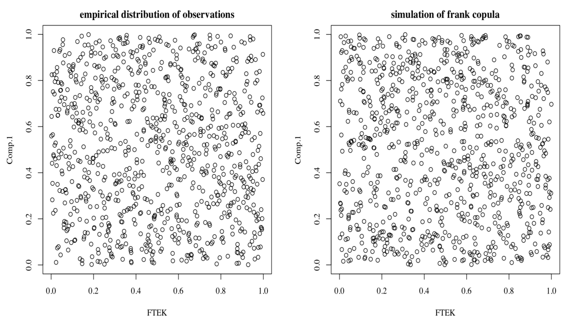
	Kendall's Tau	**Spearman's Rho**	**Max Loglikelihood**	**AIC**
Gaussian	0.3244	0.4878	171.774	-341.547
t	0.3318	0.4979	188.341	-372.682
Gumbel	0.3066	0.4399	161.841	-321.683
Clayton	0.2671	0.3895	149.611	-297.222
Frank	0.3396	0.4929	170.447	-338.894

Table 12: Fitted Copulas and Statistics

The results of the fitted copulas and statistics are summarised in Table 12. Note that Clayton copula is not applicable, as it fails to converge in maximising log-likelihood. From the table, it is easy to see that the t copula fitted the data best, with the lowest AIC of -372.682. The tail-dependence coefficients of best-fit t copula are estimated, resulting in the same lower and upper tail-dependence coefficients $\lambda_l = \lambda_u = 0.149$. The result can be visualised in Figure 7a, where both slightly lower and upper tail dependence are shown.



(a) t Copula for Whole Period data



(b) Frank Copula for stock before COVID-19

Figure 7: Simulations of Best Fitted Copula

7.3 Dependence Structure of Stock Before COVID-19

As inferred in Section 5, the sudden impact of COVID-19 starting from February 2020 has fluctuated the stock market. Thus, it arises our interest in testing whether the t copula would still be the best to fit our observations. As a result, to test this idea, we extract only part of the original log returns from January 2016 to January 2020 and implement the same procedure to conduct principal component analysis and apply the Pseudo MLE method. Upon implementation, the feature that contributes the most in Comp.2 is still FTEK, while this time, Frank copula turns out to be the best copula among the five. Since Frank copula has no lower and upper tail, FTEK and Comp.1 are asymptotically independent at tails.

	Kendall's Tau	**Spearman's Rho**	**Max Loglikelihood**	**AIC**
Gaussian	-0.037481	-0.058841	1.511830	-1.023660
t	-0.037369	-0.058665	1.074690	1.850630
Gumbel	0.000000	-0.001602	-0.000002	2.000000
Clayton	-0.030451	-0.046111	1.598480	-1.196960
Frank	-0.040406	-0.060582	1.684520	-1.369030

Table 13: Fitted Copulas and Statistics for stock before COVID-19

More obviously, as shown in Table 13, both Kendall's tau and Spearman's rho decease to a much lower level compared to the data including the COVID-19's arising period. Therefore, we can conclude that the change from the t copula to the Frank copula proves that the sudden fluctuations within the COVID-19 period have led to certain tail dependence.

8 Conclusion

As the COP26 summit is becoming increasingly popular, the release of agreements during the conference has aroused investors' insights to adapt accordingly to the transition toward a zero-carbon economy. Such initiatives would give rise to a long-term impact on the green industries, from which we can see the future trend of the green economy. Starting from the adoption of the Paris Agreement, ten green-related stocks are studied. Individual and holistic approaches are carried out through qualitative and quantitative analysis, and the hidden correlation among the stocks is discussed throughout this report. Quantitatively, a portfolio of the ten stocks selected is constructed to aware the market change. Focusing on the positive losses only, it is shown that observations in our portfolio fit the Pareto distribution. Nevertheless, when comparing the behaviour of empirical and fitted Pareto distributions, it is worth mentioning that though the Pareto distribution works reasonably well, the results may not be as accurate as the confidence level getting extremely high. On the other hand, qualitatively, the market stocks would experience periods of volatility due to changes in companies, policy reforms in sectors, as well as various market events. However, the current historical data selected does not suggest any apparent trend for the winner stocks to go upwards after the signing of the Paris Agreement. According to the COP26 summit, a clear plan on sustainability is said to be established. Thus, we expect that a more significant increase for the green stocks may be visible in 2022 or even a longer period.

Although COP26 exert a gentle influence on the green economy in general, the occurrence of some crisis events can cause shocks of the green sectors. Undoubtedly, COVID-19 is the most significant factor affecting the portfolio in recent years, which causes dramatic financial shocks in almost all industries worldwide. Such volatility appearing in the break-out period of COVID-19 also affects the dependence structure. The dependence structure of 4 bivariate stock pairs is analysed, and by implementing Principle Component Analysis, the copula that best fits the observations is obtained. Specifically, covering the whole period, t copula consistently turns out to be the best fit, considering both lower and upper tail dependence. A transition from t copula to Gaussian copula or Frank copula is shown when turning to the data before the pandemic, with the disappearance of the two-sided tail dependence. Thus, we conclude that the similar volatility within the COVID-19 period has led to some tail dependence of the data. Fitting marginal distribution as a Pareto, on the other hand, weaken the tail dependence. Therefore, we should pay more attention to marginal distributions when observing widespread climatic or social events, which, in many cases, lead to a heavy tail dependence structure.

References

1. Chrysoloras, N. (2021). An Investor's Guide to Winners and Losers of COP26 Climate Talks. Available at:
https://www.bloomberg.com/news/articles/2021-10-27/a-stock-investor-s-guide-to-the-cop26-climate-talks-in-glasgow [Accessed 29 Nov. 2021]

2. Mcwhinney, J. (2021). Top Green Investing Opportunities. Available at:
https://www.investopedia.com/articles/stocks/07/green-industries.asp [Accessed 29 Nov. 2021]

3. Luo, X. and Shevchenko, P. (2007). "The t copula with Multiple Parameters of Degrees of Freedom: Bivariate Characteristics and Application to Risk Management". *Quantitative Finance*, 10, (9). doi: 10.1080/14697680903085544

4. Stefano, D. and Alexander, J. (2005). "The t Copula and Related Copulas". *International Statistical Review*, 73, (1), pp.111-129.

5. Yamai, Y. and Yoshiba, T. (2002). "Comparative Analyses of Expected Shortfall and Value-at-Risk under Market Stress". *Monetary and Economic Studies*, 20, (3), pp.181-237.

6. IEA. (2019). "Key electricity trends". Available at:
https://www.iea.org/articles/key-electricity-trends-2018 [Accessed 3 Dec. 2021]

7. "Tesla past market value data". Available at:
http://www.xtsjzsj.com/content/if10ud3.html [Accessed 3 Dec. 2021]

8. Tidey, A. (2021). "Air pollution: Italy 'persistently' broke EU clean air laws, rules the European Court of Justice". Available at:
https://www.euronews.com/2020/11/10/air-pollution-italy-persistently-broke-eu-clean-air-laws-rules-the-european-court-of-justi [Accessed 13 Dec. 2021]

9. Fortune. (2016). "Here's Why Shares of American Airlines Are Falling Even Though It Beat Expectations". Available at:
https://fortune.com/2016/04/22/american-airlines-stock-earnings/ [Accessed 18 Dec. 2021]

10. The Motley Fool. (2016). "Why American Airlines Group, Inc. Shares Slumped 15.4% in April". Available at:
https://www.fool.com/investing/general/2016/05/08/why-american-airlines-group-inc-shares-slumped-154.aspx [Accessed 18 Dec. 2021]

11. Forbes. (2020). "American Airlines Stock Falls, But This Is What Will Come Next". Available at:
https://www.forbes.com/sites/greatspeculations/2020/10/30/american-airlines-stock-falls-but-this-is-what-will-come-next/?sh=3779b321633 [Accessed 13 Dec. 2021]

12. Forbes. (2020). "Why American Airlines Stock Has A Sizable Downside Risk Even At $9" Available at:
https://www.forbes.com/sites/greatspeculations/2020/05/07/why-american-airlines-stock-has-a-sizable-downside-risk-even-at-9/?sh=2ef3fdc97da [Accessed 13 Dec. 2021]

13. "Shipping industry responding to the COVID-19". Available at:
https://www.pwc.com/gr/en/industries/shipping-covid-19.html [Accessed 3 Dec. 2021]

14. "American Water Overview". Available at:
https://www.amwater.com/corp/About-Us/Corporate-Info/Overview [Accessed 22 Dec. 2021]

15. "About Essential Utilities". Available at:
https://www.essential.co/about [Accessed 22 Dec. 2021]

16. "About GE Renewable Energy". Available at:
https://www.ge.com/renewableenergy/about-us [Accessed 22 Dec. 2021]

17. "Corporate Profile". Available at:
https://www.ftek.com/en-US/about/company-profile [Accessed 22 Dec. 2021]

18. "Tesla's mission is to accelerate the world's transition to sustainable energy". Available at:
https://www.tesla.com/about [Accessed 22 Dec. 2021]

19. "Parent company of American Airlines". Available at:
https://www.aa.com/i18n/customer-service/about-us/american-airlines-group.jsp [Accessed 22 Dec. 2021]

20. "Profile". Available at:
https://www.eagleships.com/company/ [Accessed 22 Dec. 2021]

Appendix

Contribution Part

- 36718: (a)(b)(d)(e)
- 29393: (a)(b)(d)(e)
- 40168: (a)(b)(c)(f)
- 21851: (a)(b)(c)(f)

(a) Load Data for Portfolio

```
# read all sheets from the excel file
sheetnames <- excel_sheets("ST429_data_group6.xlsx")
mylist <- suppressMessages(lapply(excel_sheets("ST429_data_group6.xlsx"),
                                  read_excel, path = "ST429_data_group6.xlsx"))
names(mylist) <- sheetnames

# load S&P 500 data
SP500<- mylist$`S&P 500`
SP500_date<- as.Date(SP500$Date, formate= "%Y-%m-%d")
SP500<-xts(SP500$Price,order.by = SP500_date)

# load stock data
stock <- cbind(SP500_date, mylist$AWK$`Adj Close`, mylist$FTEK$`Adj Close`,
               mylist$GE$`Adj Close`, mylist$WTRG$`Adj Close`,
               mylist$TSLA$`Adj Close`, mylist$AAL$`Adj Close`,
               mylist$EGLE$`Adj Close`, mylist$ED$`Adj Close`,
               mylist$XEL$`Adj Close`, mylist$F$`Adj Close`)
colnames(stock)<-c("Date", sheetnames[-1])
stock_xts <- xts(stock[,-1], order.by = SP500_date)
```

(b) Stock Price Analysis

```
logprice<-log(stock_xts) # log return
risk_factor<-diff(logprice)[2:nrow(logprice),] # risk factor change

# log return of S&P 500 data
SP500_logprice<-log(SP500)
SP500_risk_factor <- diff(SP500_logprice)[2:nrow(SP500_logprice),]
# plot log-returns of ten stocks and S&P 500
plot.xts(merge(SP500_risk_factor, risk_factor[,1:5]), multi.panel = TRUE,
         cex = 0.2,main = "", yaxis.same = FALSE,on = NA)
```

```
plot.xts(merge(SP500_risk_factor, risk_factor[,6:10]), multi.panel = TRUE,
         cex = 0.2, main = "", yaxis.same = FALSE,on = NA)
```

(c) Loss and Risk Measure

```r
lambda <- rep(1000, 10) # 1000 shares for each stock
V0 <- sum(lambda * stock_xts[1,]) # initial value at t0
tt <- nrow(stock_xts) # current time
Vtt <- sum(lambda * stock_xts[tt,]) # current value of portfolio
weights <- lambda * stock_xts[tt,]/Vtt # weights of each stock
```

```r
# function to fit loss
loss.sim <- function(Xval, proportion, value){
  # Xval: risk factor changes
  # proportion: weights of each stock
  # value: value of portfolio
  if (is.matrix(Xval)){
    prod <- (exp(Xval)-1) %*% t(proportion)
  } else {
    n <- length(Xval)
    prod <- proportion * (exp(Xval)-1)
  }
  loss <- -value * prod
  return(loss)
}

Loss_sim <- loss.sim(risk_factor, weights, Vtt) # loss of our portfolio

Loss_positive <- Loss_sim[which(Loss_sim > 0)] # consider only positive losses
summary(Loss_positive)
hist(Loss_positive, breaks = 20, main = "" ,
     xlab = "Positive Loss", family = "Times")
```

```r
# empirical distribution
FLhat <- ecdf(Loss_positive)
plot(FLhat)
```

```r
alpha <- c(seq(0.1,0.8,0.1), seq(0.9,1,0.01)) # confidence level

# VaR and ES assuming empirical distribution
VaR_empirical <- quantile(Loss_positive, alpha, na.rm = TRUE)
ES_empirical <- rep(0, length(alpha))
for(i in 1:length(alpha)) {
  values <- Loss_positive[Loss_positive > VaR_empirical[i]]
  ES_empirical[i] <- mean(values, na.rm = TRUE)
}
```

```r
# fit pareto distribution
pareto <- suppressWarnings(fitdist(Loss_positive, "pareto"))
plot(pareto)
```

```r
alpha <- c(seq(0.1,0.8,0.1), seq(0.9,1,0.01)) # confidence level

# VaR and ES assuming pareto distribution
VaR_pareto <- qpareto(alpha, shape =pareto$estimate[1], scale =pareto$estimate[2])
```

```r
ES_pareto <- rep(0, length(alpha))
for(i in 1:length(alpha)) {
  values <- Loss_positive[Loss_positive > VaR_pareto[i]]
  ES_pareto[i] <- mean(values, na.rm = TRUE)
}
```

```r
par(family = "Times")
plot(alpha, VaR_empirical, type = "l", lty = 1, xlab = expression(alpha),
     ylab = "Estimated Risk", col = "red") # true ES_alpha
lines(alpha, ES_empirical, type = "l", lty = 1, col = "blue") # ES_alpha estimate
lines(alpha, VaR_pareto, type = "l", lty = 2, col = "red") # true ES_alpha
lines(alpha, ES_pareto, type = "l", lty = 2, col = "blue") # ES_alpha estimate

legend("topleft", bty = "n", y.intersp = 1.2, lty = c(1,1,2,2), cex = 0.8,
       col = c("red", "blue", "red", "blue"),
       legend = c(expression(VaR[alpha]^Empirical),
                  expression(ES[alpha]^Empirical),
                  expression(VaR[alpha]^Pareto),
                  expression(ES[alpha]^Pareto)))
```

```r
# for empirical distribution
# at alpha level 0.9
VaR_empirical9 <- quantile(Loss_positive, 0.9, na.rm = TRUE)
ES_empirical9 <- mean(Loss_positive[Loss_positive > VaR_empirical9])
# at alpha level 0.95
VaR_empirical95 <- quantile(Loss_positive, 0.95, na.rm = TRUE)
ES_empirical95 <- mean(Loss_positive[Loss_positive > VaR_empirical95])
# at alpha level 0.99
VaR_empirical99 <- quantile(Loss_positive, 0.99, na.rm = TRUE)
ES_empirical99 <- mean(Loss_positive[Loss_positive > VaR_empirical99])

# for pareto distribution
# at alpha level 0.9
VaR_pareto9 <- qpareto(0.9, shape = pareto$estimate[1], scale = pareto$estimate[2])
ES_pareto9  <- mean(Loss_positive[Loss_positive > VaR_pareto9])
# at alpha level 0.95
VaR_pareto95 <- qpareto(0.95, shape = pareto$estimate[1], scale = pareto$estimate[2])
ES_pareto95 <- mean(Loss_positive[Loss_positive > VaR_pareto95])
# at alpha level 0.99
VaR_pareto99 <- qpareto(0.99, shape = pareto$estimate[1], scale = pareto$estimate[2])
ES_pareto99 <- mean(Loss_positive[Loss_positive > VaR_pareto99])
```

(d) Dependence Structure

```r
# fit pairs to gaussian, t, gumbel, clayton, and frank copulas
bestfit<- function(data1,data2){
  # fix number of digits in outputs
  options(digits = 6)
  stockreturns<-merge(data1,data2)
```

```
  X <- as.matrix(stockreturns)
  # calculate losses(minus log-returns)
  X <- X[X[,1]!=0 & X[,2] !=0,]*(-1)
  # compute the empirical distribution of X
  pseudoC <<- apply(X, 2, edf, adjust = 1)
  a0<-as.numeric(c(cor(pseudoC,method = "kendall")[1,2],
                   cor(pseudoC,method = "spearman")[1,2],NA,NA))
  # set starting point
  gatheta<-cor(pseudoC,method = "spearman")[1,2]
  # fit copulas to data
  copulaXGauss <<- fitCopula(normalCopula(dim = 2,dispstr = "un"),
                             data = pseudoC,start=gatheta,method = "mpl")
  a1<-as.numeric(c(tau(normalCopula(getTheta(copulaXGauss), dim = 2)),
                   getTheta(copulaXGauss), copulaXGauss@loglik, AIC(copulaXGauss)))
  copulaXt <<- fitCopula(tCopula(dim = 2,dispstr = "un"),
                         data = pseudoC,hideWarnings=TRUE,method = "mpl")
  a2<-as.numeric(c(tau(tCopula(getTheta(copulaXt)[1], df=getTheta(copulaXt)[2],
                               df.fixed = TRUE)), getTheta(copulaXt)[1],
                   copulaXt@loglik, AIC(copulaXt)))
  gtheta <<- iTau(gumbelCopula(dim = 2), cor(pseudoC,method="kendall")[1,2])
  copulaXgumbel <<- fitCopula(gumbelCopula(dim=2), data = pseudoC,
                              start=gtheta, method = "mpl")
  a3<-as.numeric(c(tau(gumbelCopula(param = getTheta(copulaXgumbel),dim = 2)),
      rho(gumbelCopula(param = getTheta(copulaXgumbel),dim = 2)),
      copulaXgumbel@loglik, AIC(copulaXgumbel)))
  ctheta<<-iTau(claytonCopula(dim = 2),cor(pseudoC,method="kendall")[1,2])
  copulaXclayton<<- tryCatch(fitCopula(claytonCopula(dim = 2),
                                       data = pseudoC, start=ctheta,
                                       method = "mpl"), error=function(e)NA)
  a4<-as.numeric(c(tryCatch(tau(claytonCopula(param = getTheta(copulaXclayton),
                                              dim = 2)), error=function(e)NA),
                   tryCatch(rho(claytonCopula(param = getTheta(copulaXclayton),
                                              dim = 2)),error=function(e)NA),
                   tryCatch(copulaXclayton@loglik, error = function(e)NA),
                   tryCatch(AIC(copulaXclayton), error = function(e)NA)))
  ftheta <<- iTau(frankCopula(dim=2), cor(pseudoC,method = "kendall")[1,2])
  copulaXfrank <<- fitCopula(frankCopula(dim = 2), data = pseudoC,
                             start=ftheta, method = "mpl")
  a5<-as.numeric(c(tau(frankCopula(param = getTheta(copulaXfrank),dim = 2)),
                   rho(frankCopula(param = getTheta(copulaXfrank),dim = 2)),
                   copulaXfrank@loglik, AIC(copulaXfrank)))
  # return copulas with rank correlations, maximum log likelihood, and AIC
  my_matrix<-rbind(a0,a1,a2,a3,a4,a5)
  rownames(my_matrix)<-c("pseudoC","gauss","t","gumbel","clayton","frank")
  colnames(my_matrix)<-c("Tau","Rho","max. loglike","AIC")
  return(my_matrix)
}
```

```r
# fit pair XEL and ED
bestfit(risk_factor$XEL,risk_factor$ED)
# compute the lower and upper tail dependence
lambda1<-lambda(tCopula(param = getTheta(copulaXt)[1],
                        df=getTheta(copulaXt)[2],df.fixed = TRUE))
lambda1

set.seed(123)
# check the goodness of fit of the selected copula
sim1<-rCopula(nrow(pseudoC),tCopula(param =getTheta(copulaXt)[1],
                                    df=getTheta(copulaXt)[2],df.fixed = TRUE))
par(mfrow=c(1,2),cex = 0.8)
plot(pseudoC, xlab="XEL",ylab="ED",main="pseudo observations", family = "Times")
plot(sim1, xlab="XEL",ylab="ED",main="simulations", family = "Times")
```

```r
# save the copula information in global environment for future use
t1<-copulaXt
```

```r
# fit pair TSLA and F
bestfit(risk_factor$TSLA,risk_factor$F)
lambda2<-lambda(tCopula(param = getTheta(copulaXt)[1],
                        df=getTheta(copulaXt)[2],df.fixed = TRUE))
lambda2

set.seed(123)
sim2<-rCopula(nrow(pseudoC),tCopula(param = getTheta(copulaXt)[1],
                                    df=getTheta(copulaXt)[2],df.fixed = TRUE))
par(mfrow=c(1,2),cex = 0.8)
plot(pseudoC, xlab="TSLA",ylab="F",main="pseudo observations", family = "Times")
plot(sim2, xlab="TSLA",ylab="F",main="simulations", family = "Times")
```

```r
# fit pair AWK and WTRG
bestfit(risk_factor$AWK,risk_factor$WTRG)
lambda3<-lambda(tCopula(param = getTheta(copulaXt)[1],
                        df=getTheta(copulaXt)[2],df.fixed = TRUE))
lambda3

set.seed(123)
sim3<-rCopula(nrow(pseudoC),tCopula(param = getTheta(copulaXt)[1],
                                    df=getTheta(copulaXt)[2],df.fixed = TRUE))
par(mfrow=c(1,2),cex = 0.8)
plot(pseudoC, xlab="AWK",ylab="WTRG",main="pseudo observations", family = "Times")
plot(sim3, xlab="AWK",ylab="WTRG",main="simulations", family = "Times")
```

```r
# fit pair AAL and EGLE
bestfit(risk_factor$AAL,risk_factor$EGLE)
lambda4<-lambda(tCopula(param = getTheta(copulaXt)[1],
                        df=getTheta(copulaXt)[2],df.fixed = TRUE))
lambda4
```

```r
set.seed(123)
sim4<-rCopula(nrow(pseudoC),tCopula(param = getTheta(copulaXt)[1],
                                    df=getTheta(copulaXt)[2],df.fixed = TRUE))
par(mfrow=c(1,2),cex = 0.8)
plot(pseudoC, xlab="AAL",ylab="EGLE",main="pseudo observations", family = "Times")
plot(sim4, xlab="AAL",ylab="EGLE",main="simulations", family = "Times")
```

```r
# fit copula with data before June 2019
AAL_2019<-risk_factor$AAL["/2019-06"]
EGLE_2019<-risk_factor$EGLE["/2019-06"]
bestfit(AAL_2019,EGLE_2019)
```

(e) Loss and Risk Measures of Bivariate Pair

```r
# losses for pair XEL and ED
XEL.loss<-as.matrix((-1)*diff(stock_xts$XEL)[-1])
ED.loss<-as.matrix((-1)*diff(stock_xts$ED)[-1])
y<-cbind(XEL.loss,ED.loss)
# get the degree of freedom of fitted t copula
v<-getTheta(t1)[2]
# consider only positive losses
positiveloss_XEL<-XEL.loss[XEL.loss>0]
positiveloss_ED<-ED.loss[ED.loss>0]
```

```r
# fit Pareto distribution to data
marginfit_XEL<-fitdist(positiveloss_XEL,"pareto")
marginfit_ED<-fitdist(positiveloss_ED,"pareto")
# find the joint distribution model
jointcdf<-mvdc(tCopula(param = getTheta(t1)[1],dim = 2,df = v,df.fixed = TRUE),
               margins = c("pareto","pareto"),
               paramMargins = list(list(shape = marginfit_XEL$estimate[1],
                                        scale = marginfit_XEL$estimate[2]),
                                   list(shape = marginfit_ED$estimate[1],
                                        scale = marginfit_ED$estimate[2])))
```

```r
# test our joint distribution model
positivepair<-y[y[,1]>0 & y[,2]>0,]
set.seed(123)
sim_jointmodel<-rMvdc(nrow(positivepair),jointcdf)
df1<- as.data.frame(sim_jointmodel)
df2<- as.data.frame(positivepair)
ggplot(df1) + geom_point(aes(V1,V2), colour = "red") +
  xlab("XEL") + ylab("ED") +
  ggtitle("Simulations") + theme_light() +
  theme(text=element_text(family="Times"))
```

```r
ggplot(df2) + geom_point(aes(XEL,ED)) +
  xlab("XEL") + ylab("ED") +
  ggtitle("Historical Data") + theme_light() +
  theme(text=element_text(family="Times"))
```

```r
# use Monte Carlo method to estimate expected shortfall
set.seed(123)
sim_loss<-rMvdc(1000,jointcdf)
aggregateloss<-rowSums(sim_loss)
myES_t<-function(alpha){
  # compute quantiles of our simulated data
  q<-quantile(aggregateloss,probs = alpha)
  # the estimated expected shortfall converges to its true value when n is large
  mean(aggregateloss[aggregateloss>=q])
}
# vectorize the ES function so that it accepts inputs in vector
myES_t<- Vectorize(myES_t)
# expected shortfall of t copula
ES_t<-myES_t(c(seq(0.1,0.8,0.1),seq(0.9,1,0.01)))
ES_t
set.seed(123)
# expected shortfall of Gumbel copula with theta=Inf
myES_gumbel<-function(alpha){
  r1<-rpareto(1000,shape = marginfit_XEL$estimate[1],scale = marginfit_XEL$estimate[2])
  r2<-rpareto(1000,shape = marginfit_ED$estimate[1],scale = marginfit_ED$estimate[2])
  mean(r1[r1>=quantile(r1,probs = alpha)])+mean(r2[r2>=quantile(r2,probs = alpha)])
}
myES_gumbel<-Vectorize(myES_gumbel)
ES_gumbel<-myES_gumbel(c(seq(0.1,0.8,0.1),seq(0.9,1,0.01)))
ES_gumbel
loss<-rowSums(positivepair)
alpha<-c(seq(0.1,0.8,0.1),seq(0.9,1,0.01))
# compare with expected shortfall of the aggregate loss of our data
ES_data<-1:length(alpha)
for(i in 1:length(alpha)){
  ES_data[i]<-mean(loss[loss>=quantile(loss,probs = alpha[i])])
}
ES_data
```

(f) Principal Component Analysis

```r
# Calculate principal components
PCA <- prcomp(risk_factor)
summary(PCA)
PCA$rotation
PC1 <- PCA$x[,'PC1']
PC2 <- PCA$x[,'PC2']

# choose the most relevant stock in PC2
which.max(abs(PCA$rotation[,"PC2"])) # FTEK

plot(PCA, family = "Times") # plot variances
```

```
X = matrix(data = c(as.matrix(risk_factor[,'FTEK'])[,1], PC1), ncol=2)
X <- X[X[,1]!=0 & X[,2] !=0,]*(-1)
```

```
# data visualisation: plot density function
df <- data.frame(a=c(X[,1],X[,2]), Stock = c(rep('FTEK',1282),rep('PC1',1282)))
ggplot(data=df, aes(x=a, group=Stock, fill=Stock)) +
  geom_density() + theme_light() +
  theme(text=element_text(family="Times")) +
  xlim(-0.5, 0.5) + xlab(' ') + ylab('Density') +
  facet_wrap(~Stock)
```

```
# empirical distribution
ecdf1 <- ecdf(X[,1])
ecdf1 <- ecdf1(X[,1])*nrow(X)/(nrow(X)+1)
ecdf2 <- ecdf(X[,2])
ecdf2 <- ecdf2(X[,2])*nrow(X)/(nrow(X)+1)
pseudoC <- cbind(ecdf1, ecdf2)
max(pseudoC)

# perform the asymptotic independence test
BiCopIndTest(pseudoC[,1],pseudoC[,2])
```

```
# fit pairs to gaussian, t, gumbel, clayton, and frank copulas
BiCopSelect(pseudoC[,1],pseudoC[,2], familyset = c(0,1,2,3,4,5),
            selectioncrit = "AIC", indeptest = FALSE, level = 0.05, weights = NA,
            rotations = TRUE, se = FALSE, presel = TRUE, method = "mle")
a0<-as.numeric(c(cor(pseudoC,method = "kendall")[1,2],
                 cor(pseudoC,method = "spearman")[1,2],NA,NA))
```

```
# Gaussian
# set starting point
gatheta<-cor(pseudoC,method = "pearson")[1,2]
# fit copulas to data
copulaXGauss <<- fitCopula(normalCopula(dim = 2,dispstr = "un"),
                          data=pseudoC, start=gatheta,method = "mpl")
a1<-as.numeric(c(tau(normalCopula(getTheta(copulaXGauss),dim = 2)),
                 getTheta(copulaXGauss),
                 copulaXGauss@loglik,
                 AIC(copulaXGauss)))
```

```
# t copula
copulaXt <<- fitCopula(tCopula(dim = 2,dispstr = "un"),
                      data = pseudoC,hideWarnings=TRUE,method = "mpl")
a2<-as.numeric(c(tau(tCopula(getTheta(copulaXt)[1],
                             df=getTheta(copulaXt)[2],df.fixed = TRUE)),
                 getTheta(copulaXt)[1],
                 copulaXt@loglik,
                 AIC(copulaXt)))
```

```
# Gumbel
gtheta<<-iTau(gumbelCopula(dim = 2),cor(pseudoC,method="kendall")[1,2])
```

```r
copulaXgumbel<<- fitCopula(gumbelCopula(dim=2),
                           data = pseudoC, start=gtheta,method = "mpl")
a3<-as.numeric(c(tau(gumbelCopula(param = getTheta(copulaXgumbel),dim = 2)),
                 rho(gumbelCopula(param = getTheta(copulaXgumbel),dim = 2)),
                 copulaXgumbel@loglik,
                 AIC(copulaXgumbel)))
```

```r
# Clayton
ctheta<<-iTau(claytonCopula(dim = 2),cor(pseudoC,method="kendall")[1,2])
copulaXclayton<<- tryCatch(fitCopula(claytonCopula(dim = 2),
                                     data = pseudoC, start=ctheta,method = "mpl"),
                           error=function(e)NA)
a4<-as.numeric(c(tryCatch(tau(claytonCopula(param = getTheta(copulaXclayton),
                                            dim = 2)), error=function(e)NA),
                 tryCatch(rho(claytonCopula(param = getTheta(copulaXclayton),
                                            dim = 2)), error=function(e)NA),
                 tryCatch(copulaXclayton@loglik, error = function(e)NA),
                 tryCatch(AIC(copulaXclayton), error = function(e)NA)))
ftheta<<- iTau(frankCopula(dim=2), cor(pseudoC, method = "kendall")[1,2])
```

```r
# Frank
copulaXfrank<<- fitCopula(frankCopula(dim = 2),
                          data = pseudoC, start=ftheta, method = "mpl")
a5<-as.numeric(c(tau(frankCopula(param = getTheta(copulaXfrank), dim = 2)),
                 rho(frankCopula(param = getTheta(copulaXfrank), dim = 2)),
                 copulaXfrank@loglik,
                 AIC(copulaXfrank)))
```

```r
# copulas with rank correlations, maximum log likelihood, and AIC
my_matrix<-rbind(a0,a1,a2,a3,a4,a5)
rownames(my_matrix)<-c("pseudoC","gauss","t","gumbel","clayton","frank")
colnames(my_matrix)<-c("Tau","Rho","max. loglikelihood","AIC")
my_matrix
```

```r
# Simulate the best fitted t copula
lambda1<-lambda(tCopula(param = getTheta(copulaXt)[1],
                        df=getTheta(copulaXt)[2],df.fixed = TRUE))
sim1<-rCopula(nrow(pseudoC),tCopula(param = getTheta(copulaXt)[1],
                                    df=getTheta(copulaXt)[2],df.fixed = TRUE))
par(mfrow=c(1,2))
plot(pseudoC, xlab="FTEK",ylab="Comp.1",
     main="empirical distribution \nof observations")
plot(sim1, xlab="FTEK",ylab="Comp.1",
     main="simulation of \nt copula")
```

```r
# For data before COVID-19, change data by:
# stocks = stocks['2016-01-04/2020-01']
# redo the above to fit copulas again, the best fitted copula changes to Frank Copula
# simulate the best fitted Frank copula
```

```
# getTheta(copulaXfrank)
# lambda1<-lambda(frankCopula((param = getTheta(copulaXfrank))))
# sim1<-rCopula(nrow(pseudoC),frankCopula(param = getTheta(copulaXfrank)))
# par(mfrow=c(1,2))
# plot(pseudoC, xlab="GE",ylab="Comp.1",
# main="empirical distribution \nof observations")
# plot(sim1, xlab="GE",ylab="Comp.1",
# main="simulation of \nfrank copula")
```