VRIJE UNIVERSITEIT

Group Assignment 3: Multivariate Analysis with PCA/FA/Copula/EVT

Quantitative Financial Risk Management

Authors: A 15

Jip de Boer j.a4.de.boer@student.vu.nl (2710065)

Yunji Eo y.eo@student.vu.nl (2735445)

Lars de Graaff l.de.graaff@student.vu.nl (2790679)

Xuan Zhou x.zhou9@student.vu.nl (2653511)

Supervisor:

Dr. C.S. Bos

May 15, 2025

1 Introduction

This report aims to analyze the risk and dependence structure of a diversified portfolio using advanced multivariate techniques, including Principal Component Analysis (PCA), Factor Analysis (FA), Copulas and Extreme Value Theory (EVT). Our analysis is carried out on a portfolio consisting of 10 assets from different sectors denominated in five currencies: USD, EUR, JPY, HKD and CNY. To provide a general overview, the different currency types, the components and their description are displayed in Table 1 below.

Table 1: Asset Composition: Equity constituents grouped by currency and company sector.

Currency	**Components**	**Company Sector / Description**
USD	Apple (AAPL), Microsoft (MSFT)	U.S. technology giants providing global exposure to hardware, software and cloud services
EUR	SAP (SAP.DE), Adidas (ADS.DE)	German firms in enterprise software and consumer goods (apparel and sportswear)
JPY	Sony (6758.T), SoftBank Group (9984.T)	Japanese exposure to consumer electronics, media and investment holdings
HKD	Tencent (0700.HK), HSBC Holdings (0005.HK)	Hong Kong-based leaders in internet services and global banking
CNY	Ping An Insurance (601318.SS), Kweichow Moutai (600519.SS)	Chinese companies in financial services and premium consumer beverages

This composition ensures sensitivity to stock market movements, regional economic conditions and exchange rate fluctuations. The primary goal is to assess the portfolio's risk profile and dependence structure using multivariate techniques, providing insights into diversification benefits and tail risk.

The report proceeds as follows. Section 2 provides a preliminary analysis of the data at hand from the selected assets. Section 3 presents Principal Component Analysis (PCA) to identify key factors driving portfolio variability. Section 4 performs Factor Analysis (FA) using both macroeconomic factors and statistical factors derived from PCA. Section 5 introduces the concept of copulas to capture non-linear dependencies and tail risk between asset pairs. Finally, Section 6 uses Extreme Value Theory (EVT) to model extreme losses and assess tail risk from a different perspective. To conclude, a final verdict on the lessons learned from our risk and dependence analysis is presented in Section 7.

2 Data

This section provides a preliminary analysis of the data at hand from the selected assets. To this end, we discuss the historical return characteristics of the selected equity components by examining descriptive statistics, stock price evolution and volatility, return time series and their volatility, as well as the distributional properties of returns.

Table 2 provides descriptive statistics for the ten equity assets from January 1, 2015 to December 31, 2024. The U.S. tech stocks (AAPL, MSFT) show the highest mean returns (0.096%) with moderate volatility, indicating strong risk-adjusted performance. In contrast, 0005.HK has the lowest mean (0.023%) and lowest volatility. Volatility is highest for 9984.T and 0700.HK, reflecting greater return variability. Most assets display mild skewness, except SAP.DE, which is strongly negatively skewed (-1.565), suggesting downside risk. All returns are leptokurtic, with SAP.DE exhibiting extreme kurtosis (30.069), indicating fat tails and a higher probability of extreme returns.

Table 2: Descriptive statistics of daily continuously compounded returns for the 10 equity assets, spanning January 2015 to December 2024. The table reports the mean, standard deviation, quantiles and higher moments (skewness and kurtosis) for each asset across different currencies and sectors.

Statistic	0005.HK	0700.HK	600519.SS	601318.SS	6758.T	9984.T	AAPL	ADS.DE	MSFT	SAP.DE
Count	2466	2466	2466	2466	2466	2466	2466	2466	2466	2466
Mean	0.023	0.058	0.093	0.027	0.079	0.041	0.096	0.064	0.096	0.065
Std Dev	1.425	2.194	1.909	1.835	1.961	2.498	1.777	1.983	1.706	1.560
Min	-9.996	-13.180	-15.213	-10.540	-14.723	-20.651	-13.771	-14.394	-15.945	-24.766
25%	-0.647	-1.134	-0.894	-0.913	-0.981	-1.122	-0.727	-0.908	-0.643	-0.664
Median	0.000	0.000	0.000	-0.023	0.039	0.000	0.092	0.024	0.090	0.078
75%	0.726	1.210	1.025	0.833	1.066	1.256	0.997	1.010	0.936	0.848
Max	8.819	20.827	9.531	9.545	11.695	17.355	11.316	19.379	13.293	11.822
Skewness	-0.227	0.318	0.081	0.223	-0.003	-0.298	-0.163	0.430	-0.156	-1.565
Kurtosis	5.250	6.285	4.248	4.259	5.531	8.794	5.681	8.471	8.348	30.069

Additional exploratory figures supporting the analysis can be found in Appendix A.2. Figure 4 presents the stock price evolution of the ten selected equities over the sample period (1 Jan. 2015 - 31 Dec. 2024). Each asset displays long-term growth trends in its respective local currency, while also capturing major market events such as the COVID-19 crash and subsequent recovery. The visual patterns reveal periods of elevated volatility which underscores the importance of risk modeling techniques that can adapt to such regime changes.

The daily return plots in Figure 5 further illustrate these dynamics. From these return plots it can be observed that there appear to be clusters of volatility indicated by changing volatility for different periods in time, for example indicated by the higher volatility around the start of COVID (from 2020 onwards). This further supports the hypothesis that regime shifts occur over time and should ideally be accounted for by the risk modeling techniques. At the very least, analysts should be aware of these patterns to make more informed decisions about which methods are most suitable for accurately analyzing risk, given the regime that is present in the data at the time of analysis.

Lastly, Figure 7 and 8 show histograms of the daily log returns for the 10 asset constituents with fitted Gaussian and Student's t-distributions. While the return distributions approximate a bell curve, the presence of heavy tails, suggests deviations from normality. This observation is critical for risk analysts, as models based on the normal distribution may underestimate the probability and magnitude of extreme return events, both negative and positive. And indeed, when fitting Student's t-distributions to the assets returns, we observe a better distributional fit as indicated by the histogram bars being closely aligned with the Student-t's distribution curve in Figure 8.

3 PCA

To obtain the principal components (PC's) we base the analysis on the covariance matrix of the demeaned return series, rather than the correlation matrix. This approach ensures that assets with higher variance have a proportionally greater influence on the extracted components. Using the covariance matrix is appropriate here because all return series are expressed in the same units (percentage daily returns) and have not been standardized. For a reference of the mathematics behind the construction of the principal components, please see Appendix A.1.

The eigenvalues and variance decomposition shown in Table 3 help answer how many factors are needed to summarize the variability in the return series. We investigated both the Kaiser criterion (include all PC's of which the corresponding eigenvalue ≥ 1) and the cumulative variance threshold of 90%. By doing so, we find that, given these rules, it is optimal to retain the 7 principal components (PC's) which together

Table 3: Principal Component Analysis results: this table reports the eigenvalues, individual explained variance and cumulative explained variance for all ten principal components. Values are rounded to two decimal places, which may cause the cumulative explained variance to slightly exceed 100%.

Component	Eigenvalue	Explained Variance	Cumulative Variance
PC1	11.38	31.27%	31.27%
PC2	5.43	14.94%	46.21%
PC3	5.17	14.20%	60.41%
PC4	3.35	9.20%	69.61%
PC5	2.78	7.65%	77.26%
PC6	2.55	7.02%	84.28%
PC7	**1.85**	**5.09%**	**89.37%**
PC8	1.53	4.22%	93.59%
PC9	1.41	3.88%	97.47%
PC10	0.92	2.54%	100.00%

explain 89.73% of the variance. Also all of these PC's have a corresponding eigenvalue ≥ 1. However, not all PC's with a corresponding eigenvalue $\geq$1 are included as this would result in including 9 PC's which still would contain almost all dimensions where we actually want to use PCA as a dimension reduction method for our data set. It is furthermore interesting to see that for 5 PC's already more than 75% of the variance in the data of these 10 assets is explained. This indicates that a relatively small number of factors (PC's in this case) can be sufficient to summarize most of the information in the asset returns. Also, the higher numbered PC's appear to have diminishing marginal contributions to explaining the variance in the data indicated by the decrease in Explained Variance with respect to the ordered PC's, which are ordered based on their corresponding eigenvalues from large to small. Note for example that the first PC explains more variance present within the data than the last 6 PC's together.

Table 4 displays the factor loadings matrix which shows how each asset contributes to the retained PC's. The first few PC's appear to capture interpretable sources of common variation. For example, PC1 is broadly negative among all assets, which likely indicates a common market wide movement. PC2 and PC3 display more dissimilarities among the loading indicated by both of them having negative and positive loadings. Specifically, U.S. tech stocks (AAPL, MSFT) load positively on PC2, while certain Asian stocks such as SoftBank (9984.T), Sony (6758.T), Tencent (0700.HK) and Ping An Insurance (601318.SS) load negatively. The loading patterns thus seem to reflect broad regional or sectoral groupings. This

Table 4: Factor loadings matrix from Principal Component Analysis: each entry shows the loading (correlation-like weight) of a specific asset on a principal component (PC1 to PC9). These loadings indicate the contribution of each asset to the corresponding component. Values are rounded to two decimal places.

Asset	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
0005.HK	-0.20	-0.07	0.02	0.07	-0.06	-0.10	0.61	0.06	0.75
0700.HK	-0.44	-0.08	0.40	0.06	-0.74	0.25	-0.05	-0.01	-0.13
600519.SS	-0.30	-0.06	0.50	-0.13	0.40	-0.02	-0.57	0.09	0.39
601318.SS	-0.28	-0.08	0.44	-0.14	0.39	-0.18	0.51	-0.11	-0.48
6758.T	-0.34	-0.19	-0.28	-0.04	-0.19	-0.83	-0.19	-0.01	-0.06
9984.T	-0.53	-0.40	-0.53	-0.13	0.22	0.45	-0.01	-0.05	-0.06
AAPL	-0.20	0.54	-0.11	-0.42	-0.05	-0.01	-0.02	-0.24	0.03
ADS.DE	-0.28	0.35	-0.06	0.77	0.18	0.01	-0.07	-0.40	-0.01
MSFT	-0.19	0.53	-0.11	-0.34	-0.02	0.05	0.04	-0.01	0.03
SAP.DE	-0.22	0.29	-0.09	0.22	0.08	0.00	0.04	0.87	-0.15

suggests the first few components capture economically meaningful variation. Negative loadings are not unexpected and can for example imply an inverse relationship between the asset's movement and that of the corresponding components.

Table 5: Explained variance in asset returns by the retained principal components. This table shows the R^2 for each asset, indicating how well the factors explain its return variation. Values are rounded to two decimal places.

0005.HK	0700.HK	600519.SS	601318.SS	6758.T	9984.T	AAPL	ADS.DE	MSFT	SAP.DE
1.00	1.00	1.00	1.00	1.00	1.00	0.88	1.00	0.83	0.99

To asses how well the selected PC's explain individual asset returns we regress each asset's return on the full set of the 7 selected PC's. The resulting R^2 values are reported in Table 5. These R^2 values show that nearly all assets are almost perfectly explained by the selected factors. Most assets have R^2 values close to 1.00, with only AAPL and MSFT slightly lower at 0.88 and 0.83 respectively. This confirms that the PCA model captures nearly all relevant variability in the data which strengthens the decision of the number of factors that are selected using the 90% variance threshold.

4 Factor Analysis

For the process of the data dynamic, factor analysis notation is as follows,

$$\mathbf{X}_t = \mathbf{a} + \mathbf{B}\mathbf{F}_t + \boldsymbol{\epsilon}_t$$

where $\mathbf{F_t}$ denotes the vector of common factors, $\boldsymbol{\epsilon}_t$ denotes the vector of idiosyncratic error terms, $\mathbf{B}$ is the matrix of constant factor loading and $\mathbf{a}$ is the vector of constants.

In this report, we apply the two methods of factor analysis; the first is to introduce macro factors and the second is to include PCA factors as inputs to explain the returns of the sample assets. The macro-factors included are as follows; all data have the same time range as the original asset data from January 1, 2015 to December 31, 2024.

- Daily Fama-French 5 factors and momentum for developed markets, North America, Japan, Europe and Asia-Pacific countries excluding Japan with the factors of Mkt-RF (Market excess return), SMB (Small Minus Big), HML (High Minus Low), RMW (Robust Minus Weak), CMA (Conservative Minus Aggressive), RF (Risk free rate), WML (momentum), where RF is used to calculate excess returns. And data is downloaded from the website of French, 2024.
- VIX denotes the CBOE volatility index based on S&P 500 to proxy the general volatility of the market.
- For the analysis of individual countries, the daily currencies' exchange rate towards United State Dollars (USD) are applied, including exchange rate for People's Republic of China (Yuan/US$), exchange rate for Hong Kong (Dollar/US$), exchange rate for Japan (Yen/US$), exchange rate for European Monetary Union (Euro/US$).

Macro-factor analysis - overall

For the overall analysis, we use Fama-French 5-factor data collected from developed countries and CBOE volatility index of VIX into the regression. In Table 6, the results for the estimates of **B** of macro-factors and corresponding regression R^2 are reported, where the former shows the explainatory degrees coming from the factor and the latter measures the regression relationship between X_j and F for j-th assets and can be interpreted by the percentages of variations that can be explained by the regression model.

We observe the following findings from the results, for the factor of Mkt-RF, all assets obtain positive and statistically significant coefficients, indicating that the positive relationship with the general market portfolio returns, especially for APPL (Apple) and 9984.T (SoftBank Group). And for the factor of CMA, all assets also show statistically significant coefficients while with mixed signs, where 6758.T (Sony), 9984.T (SoftBank Group) and MSFT (Microsoft) are positive. Besides, the coefficients of the volatility proxy VIX are quite small compared to other involved factors even close to zero, the responses towards this factor deviations are limited. Finally, for (adjusted) R^2, assets in US stock exchange markets obtain higher R^2 values, following those in EU markets and those listed in Chinese markets get the lowest results, which can be explained by the smaller proportions coming from the Fama-French 5 factors developed countries dataset and the less relevant to the selected volatility index.

Table 6: Macro factor Analysis - multivariate regression results

Asset	const	Mkt-RF	SMB	HML	RMW	CMA	WML	VIX	R^2	adj R^2
0005.HK	0.25*	0.50*	0.66*	0.55	0.19	-0.37*	-0.04	-0.01*	15.75%	15.51%
0700.HK	0.07	0.72*	1.09*	-0.24	-0.03	-0.47*	0.03	0.00	11.90%	11.65%
600519.SS	0.11	0.42*	0.59*	-0.02	0.04	-0.27	0.13*	0.00	4.68%	4.41%
601318.SS	0.24*	0.45*	0.60*	0.11	0.13	-0.20	0.04	-0.01*	5.77%	5.50%
6758.T	0.12	0.78*	1.26*	0.26*	0.75*	0.44*	0.24*	0.00	13.69%	13.44%
9984.T	0.29*	1.02*	1.80*	0.28	0.80*	0.62*	0.24*	-0.02*	15.97%	15.73%
AAPL	0.12	1.18*	-0.76*	-0.74*	0.65*	-0.35*	0.06	-0.01	56.31%	56.18%
ADS.DE	0.14	0.93*	0.53*	0.03	0.50*	-0.41*	-0.19*	-0.01	20.36%	20.13%
MSFT	0.06	1.04*	-1.19*	-0.34*	0.59*	0.57*	0.11*	-0.01	67.25%	67.16%
SAP.DE	0.09	0.85*	0.11	-0.27*	0.26*	-0.04	0.00	0.00	27.47%	27.26%

p-value$<$0.05 means statistically significant at the 95% confidence level, denoted with *.

The residual error correlation matrix of the assets is also reported in Table 7, where results greater than 0.2 are indicated by the **bold** font. This matrix can be used to measure whether the unexplained parts of the models (residual errors) are still correlated. The results show that there exist country-wise factors that are not captured by the current models and that the assets in China and Hong Kong are relatively highly correlated. For Chinese assets, the correlation between 600519.SS and 601318.SS is close to 0.5 and 0700.HK is correlated to two Chinese assets with a magnitude of close to 0.3. The residuals coming from APPL and MSFT, ADS.DE and SAP.DE are also weakly correlated.

Table 7: Residual error correlation matrix of asset pairs - macro factor

	0005.HK	0700.HK	600519.SS	601318.SS	6758.T	9984.T	AAPL	ADS.DE	MSFT	SAP.DE
0005.HK	1.000									
0700.HK	**0.252**	1.000								
600519.SS	0.123	**0.353**	1.000							
601318.SS	**0.221**	**0.311**	**0.494**	1.000						
6758.T	0.181	0.153	0.063	0.085	1.000					
9984.T	0.191	0.178	0.104	0.090	**0.349**	1.000				
AAPL	-0.042	-0.030	0.002	-0.022	-0.021	-0.049	1.000			
ADS.DE	0.061	0.099	0.072	0.031	0.033	0.037	-0.103	1.000		
MSFT	-0.012	-0.054	-0.023	-0.036	-0.083	-0.056	**0.214**	-0.101	1.000	
SAP.DE	0.062	0.058	0.049	0.025	0.046	0.067	-0.115	**0.278**	-0.054	1.000

Macro-factor analysis - by region

Based on the findings of the overall report multivariate regression results, we further split the assets by region and perform the regressions for comparison. Differently, for the applied Fama-French 5 factor data, we choose regional specific ones to improve the model accuracy (North America dataset for United States assets, Japan datasets for Japanese assets, Europe datasets for two Germany assets and Asia-Pacific countries excluding Japan datasets for Hong Kong and Chinese assets) and also introduce the daily currencies' exchange rate towards United States Dollars (USD) to indicate the macro factor influences. The results of the analysis are shown in Table 8. First, the overall values of all countries R^2 increase compared to the previous results, indicating the explanatory efficiency of introducing country-specific datasets. Second, in this case, more factors that come from Fama-French become statistically significant, though the momentum factor is only significant for parts of United States, Hong Kong and Chinese assets. Third, for the country-specific factor, in the case of US assets, we still use the volatility proxy factor of VIX in regression, which has a coefficient of zero and is not statistically significant. For the rest of the country cases with the introduction of the currency exchange rate, only for SAP.DE assets in Germany shows the positive and statistically significant coefficient, where the value of R^2 increases from 27.47% to 40.02%, this finding indicates that there are still unobserved factors that should better explain asset returns.

Table 8: Macro factor analysis - multivariate regression results by regions

Asset	const	Mkt-RF	SMB	HML	RMW	CMA	WML	VIX	Euro/US	JP/US	HK/US	Yuan/US	R^2
AAPL	0.08	1.15*	-0.23*	-0.55*	0.60*	0.29*	0.05	0.00					61.35%
MSFT	0.00	1.15*	-0.52*	-0.16*	0.45*	0.45*	0.11*	0.00					70.94%
SAP.DE	-1.47*	0.75*	-0.68*	-0.53*	-0.22	0.33*	0.01		1.69*				40.02%
ADS.DE	0.48	0.93*	-0.42*	-0.15	0.43*	0.39*	-0.09		-0.51				32.91%
6758.T	0.02	0.80*	-0.65*	-0.22*	0.18	0.61*	0.09			0.00			35.17%
9984.T	-0.41	0.92*	-0.48*	0.41*	0.87*	-1.73*	0.02			0.00			34.40%
0700.HK	10.46	1.24*	-0.14	-0.45*	-1.05*	-0.16	0.02				-1.33		30.70%
0005.HK	-5.07	1.02*	-0.04	0.72*	-0.16*	-0.22*	-0.14*				0.65		37.34%
601318.SS	-0.85	1.06*	0.34*	0.41*	-0.38*	0.01	-0.14*					0.13	20.65%
600519.SS	0.42	0.79*	0.32*	-0.14	-0.53*	-0.01	-0.03					-0.05	12.78%

p-value$<$0.05 means statistically significant at the 95% confidence level, denoted with *.

Statistical factor analysis - PCA

In the following, we apply the alternative method to take the PCA rotated factors as regression variables, which is also known as the statistic method for estimation, where both $\mathbf{B}$ and $\mathbf{F}_t$ are unknown compared to the macro method mentioned above with the only unknown $\mathbf{B}$. We apply the results of section 3 and take the 7 principal components into account in the analysis; the results of the multivariate regression are shown in Table 9 and the corresponding residual error correlation matrix is listed in Table 10.

Table 9: Statistical factor analysis - multivariate regression results from PCA

Asset	const	PC1	PC2	PC3	PC4	PC5	PC6	PC7	R^2	adj R^2
0005.HK	0.02	0.10*	-0.04*	0.05*	0.10*	0.16*	-0.06*	0.61*	60.52%	60.41%
0700.HK	0.06*	0.04*	0.01*	0.12*	0.08*	0.97*	0.03*	-0.12*	99.49%	99.49%
600519.SS	0.09*	0.01	-0.03*	0.75*	-0.04*	-0.10*	0.03*	-0.51*	93.81%	93.80%
601318.SS	0.03*	0.00	-0.04*	0.64*	0.05*	-0.07*	0.04*	0.57*	89.65%	89.62%
6758.T	0.08	0.13*	-0.04*	0.04*	0.04*	-0.01*	-0.98*	-0.04*	99.80%	99.80%
9984.T	0.04*	0.98*	-0.04*	-0.03*	-0.02*	-0.05*	0.14*	-0.06*	99.86%	99.86%
AAPL	0.10*	0.03*	0.71*	0.04*	-0.06*	0.00	-0.05*	-0.01	84.75%	84.70%
ADS.DE	0.06*	-0.01*	-0.04*	0.04*	0.91*	-0.10*	-0.02*	-0.07*	93.70%	93.68%
MSFT	0.10*	0.03*	0.67*	0.02*	0.00	-0.01	0.01	0.04*	82.59%	82.54%
SAP.DE	0.07*	0.05*	0.19*	0.03*	0.39*	-0.05*	-0.02	0.04*	50.07%	49.93%

p-value$<$0.05 means statistically significant at the 95% confidence level, denoted with *.

The results in Table 9 provide us with the following observations, first, compared to the previous two cases of general and by region, the (adjusted) values of R^2 increase greatly with the magnitudes of 80% for the majority of investigated assets, except for 0005.HK (HSBC Holdings) and SAP.DE. Secondly, almost all assets obtain the statistically significant coefficients for the factors, except MSFT (Microsoft) has PC4, PC5, PC6 remains insignificant. Third, component-specific related features can be detected. For example, PC1 and PC6 have an absolute value of 0.98 for 9984.T (Softbank Group) and 6758.T (Sony), respectively, which brings the country-wise indicator. PC3 has significantly higher coefficients for Chinese assets compared to the rest. And APPL (Apple) and MSFT (Microsoft) have the value of 0.71 and 0.67 for PC3, which may also indicate features of the technology industry.

However, when checking the residual error correlation matrix based on multivariate regression on PCA factors, as shown in Table 10, the observed correlations have significantly higher values compared to the previous results of the macro method. Specifically, we observe that for assets in Asian regions (Hong Kong, Japan and China), the residual error correlations are larger than 0.7, where for Hong Kong and China assets the values reach 0.98, also the absolute value 0.99 for asset pair for Hong Kong and China, indicating the current factor setting leaves features among Asian regions undiscovered. Similarly, a high correlation between SAP.DE and ADS.DE with a value of -0.98 is also observed. And the strong and middle level correlations between APPL and MSFT, APPL and 6758.T (Sony) are also noted, which may bring the concerns of imperfections especially for the technology stocks.

Table 10: Residual error correlation matrix of asset pairs - PCA factor

	0005.HK	0700.HK	600519.SS	601318.SS	6758.T	9984.T	AAPL	ADS.DE	MSFT	SAP.DE
0005.HK	1.00									
0700.HK	**-1.00**	1.00								
600519.SS	**0.99**	**-0.98**	1.00							
601318.SS	**-0.99**	**0.99**	**-1.00**	1.00						
6758.T	**-0.85**	**0.85**	**-0.83**	**0.84**	1.00					
9984.T	**-0.82**	**0.81**	**-0.90**	**0.90**	**0.74**	1.00				
AAPL	0.05	-0.06	-0.06	0.03	**-0.47**	**-0.23**	1.00			
ADS.DE	-0.11	0.09	**-0.26**	**0.26**	**0.21**	**0.65**	**0.37**	1.00		
MSFT	0.02	-0.02	0.07	-0.04	**0.50**	0.04	**-0.89**	0.08	1.00	
SAP.DE	-0.08	0.10	0.07	-0.07	-0.09	-0.05	**-0.30**	**-0.98**	-0.16	1.00

Comparison between macro-factors and statistical factors

Finally, for the factor analysis part analysis, we include the overall correlation matrix of the macro factors that were applied for overall multivariate regressions and the principal component factors based on the PCA statistical method, shown in Table 11. We observe that within principal components, the strong correlations are shown for PC1 with PC3-6, between PC2 and PC4, also for PC3 and PC4-5, which shows for such pairs that they might be capable of capture similar features. And for PC7, which are relatively less correlated with others. For the macro-factor domain, notably, a strong positive correlation is observed for CMA and HML with the correlation value of 0.76, economically meaning that the value stocks tend to invest conservatively, it also brings potential impacts of multicollinearity towards the current model setting.

For the relations between principal components and macro factors, Mkt-RF shows strong correlations with PC1-5, especially for PC2, the value reaches 0.73, indicating that PC2 may capture the general risks of the market.

Table 11: Correlation matrix between PCs, Fama-French Factors and volatility proxy

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	Mkt-RF	SMB	HML	RMW	CMA	RF	WML	VIX
PC1	1.00														
PC2	0.11	1.00													
PC3	**0.28**	0.12	1.00												
PC4	**0.24**	**0.28**	**0.23**	1.00											
PC5	**0.27**	0.07	**0.42**	0.19	1.00										
PC6	**-0.38**	-0.11	-0.15	-0.18	-0.17	1.00									
PC7	0.05	0.01	0.06	0.02	0.01	-0.07	1.00								
Mkt-RF	**0.35**	**0.73**	**0.28**	**0.48**	**0.20**	-0.27	0.12	1.00							
SMB	0.12	-0.57	0.03	-0.08	0.09	-0.07	0.02	-0.34	1.00						
HML	-0.04	-0.39	-0.06	-0.08	-0.07	0.03	0.16	-0.09	**0.22**	1.00					
RMW	-0.05	0.15	-0.06	-0.05	-0.08	0.01	-0.07	-0.16	-0.24	-0.34	1.00				
CMA	-0.16	-0.44	-0.15	-0.22	-0.15	0.11	0.07	-0.33	0.16	**0.76**	-0.07	1.00			
RF	0.00	0.00	-0.02	0.01	-0.01	0.01	0.04	-0.02	-0.05	-0.02	-0.01	-0.01	1.00		
WML	-0.02	0.06	-0.02	-0.14	-0.03	0.00	-0.12	-0.16	-0.10	-0.35	0.17	-0.12	0.02	1.00	
VIX	-0.13	-0.12	-0.09	-0.11	-0.05	0.07	-0.09	-0.18	0.02	-0.02	0.04	0.06	-0.10	-0.01	1.00

5 Copulas

In this section, we use the concept of copulas to model and analyze the dependence structure between different pairs of asset returns. From the initial set of 10 assets used throughout our report, we select four representative pairs for detailed analysis. The pairs are as follows: AAPL-MSFT (USD), SAP.DE-ADS.DE (EUR), 0700.HK-0005.HK (HKD), 6758.T-9984.T (JPY). Each pair is chosen based on the same currency region. This helps ensure that the dependency structure is not distorted by currency uncertainty.

Since we examine pairs of assets, we deal with bivariate data. Let us therefore consider that each pair of asset returns can be decomposed into two real-valued random variables, X_1 and X_2, where X_1 represents the returns of asset 1 and X_2 represents the returns of asset 2. To give a general sense of the relationship between each pair of variables, Figure 10 in Appendix A.2 presents the bivariate scatter plots of daily log returns for the four selected asset pairs. These plots are enhanced with marginal histograms and colored using kernel density estimation (KDE). While quick and intuitive, these standard scatter plots are limited to visual pattern recognition. Differences in scale, distributions and outliers can distort the dependence structure between variables, as seen in Figure 10.

In contrast, copulas explicitly capture the dependence structure between variables independently of their marginal distributions. By definition, a d-dimensional copula is a multivariate distribution function $C : [0,1]^d \to [0,1]$ with standard uniform marginal distributions (in our report, we use d=2 for the dimension, as we are working with bivariate data). By transforming each variable to its rank and mapping it to the $[0,1]$ range, copulas reveal how variables co-move based on their joint behavior, unaffected by scale or distribution. In other words, if we choose a copula and some marginal distributions and entangle them in the right way, we will end up with a proper multivariate distribution function. This is due to Sklar's theorem: Let F be a joint cumulative distribution function with marginals $F_1, \ldots, F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that, for all $\mathbf{x} \in \mathbb{R}^d$, $F(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots, F_d(x_d)\right)$. According to Sklar's Theorem, the unique copula C associated with a joint distribution function F and marginal distribution functions $F_1, \ldots, F_d$ is given by: $C(u_1, \ldots, u_d) = F\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right)$.

We consider four different families of copulas: Gaussian, Student's t, Clayton and Gumbel copulas. These copulas capture a variety of dependence structures, including symmetric and asymmetric tail dependencies. The general multivariate forms of these copulas are as follows:

- **Gaussian copula:** captures symmetric dependence without tail dependence.

  $$C_{\boldsymbol{P}}^{\mathrm{Ga}}(u_1, \ldots, u_d) = \Phi_{\boldsymbol{P}}\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right),$$

  where Φ^{-1} is the inverse standard normal CDF and $\Phi_{\boldsymbol{P}}$ is the CDF of a d-dimensional standard normal distribution with correlation matrix $\boldsymbol{P}$.

- **Student's t copula:** allows for symmetric dependence and captures tail dependence in both tails.

  $$C_{\nu,\boldsymbol{P}}^{t}(u_1, \ldots, u_d) = t_{\nu,\boldsymbol{P}}\left(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)\right),$$

  where t_ν^{-1} is the inverse univariate t CDF with ν degrees of freedom and $t_{\nu,\boldsymbol{P}}$ is the CDF of the multivariate t-distribution.

- **Clayton copula:** an Archimedean copula that exhibits lower tail dependence.

$$C_\theta^{\mathrm{Cl}}(u_1, \ldots, u_d) = \left(\sum_{i=1}^{d} u_i^{-\theta} - d + 1 \right)^{-1/\theta}, \quad \theta > 0.$$

- **Gumbel copula:** an Archimedean copula that exhibits upper tail dependence.

$$C_\theta^{\mathrm{Gu}}(u_1, \ldots, u_d) = \exp\left\{ - \left[\sum_{i=1}^{d} (-\log u_i)^\theta \right]^{1/\theta} \right\}, \quad \theta \geq 1.$$

The four families of copulas are fitted using canonical maximum likelihood estimation (CMLE). To compare the goodness-of-fit of the different copula models, the Akaike Information Criterion (AIC) is used. It is defined as:

$$\mathrm{AIC} = 2k - 2LL,$$

where k is the number of estimated parameters and LL is the maximized log-likelihood of the model. For each pair of assets, we select the copula family with the lowest AIC as the best-fitting model. The copula estimation results for each asset pair, can be found in Tables 12-15. For all four asset pairs, the Student-t copula consistently shows the highest log-likelihood (LL) and the lowest AIC, indicating that it is most appropriate for modeling joint dependencies between these equity pairs, capturing both correlation and tail risk more effectively than the other copula types.

Visual representations of the copulas for each asset pair are presented in Figures 11-18. The first four figures illustrate the copula-based dependence structures simulated from the empirical data for each asset pair. Each panel visualizes how the respective copula captures the dependence present in the observed data. The subsequent four figures show simulated copula samples on the uniform scale [0,1], where no marginal transformation has been applied. These plots provide a standardized view of the dependence structure, isolated from the marginal distributions. Notably, the visualizations already reveal clear indications of symmetric dependence as well as upper and lower tail dependence, particularly in the cases of the Gumbel and Clayton copulas. This represents a significant improvement over the earlier scatter plots in Figure 10, which offered a more limited view of the dependence structure and lacked clarity in the tails.

Table 12: AAPL-MSFT Copula Estimation.

	Gaussian/Student-t		Gumbel/Clayton	
Method	C_P^{Ga}	$C_{\nu,P}^t$	C_θ^{Gu}	C_θ^{Cl}
ρ	0.668	0.676	–	–
ν	–	3.011	–	–
θ	–	–	1.850	1.380
LL	724.805	850.267	732.999	675.201
AIC	-1447.610	**-1696.534**	-1463.999	-1348.402

Table 13: SAP.DE-ADS.DE Copula Estimation.

	Gaussian/Student-t		Gumbel/Clayton	
Method	C_P^{Ga}	$C_{\nu,P}^t$	C_θ^{Gu}	C_θ^{Cl}
ρ	0.490	0.513	–	–
ν	–	4.995	–	–
θ	–	–	1.465	0.774
LL	335.975	387.852	325.369	309.989
AIC	-669.951	**-771.703**	-648.738	-617.978

Table 14: 0700.HK-0005.HK Copula Estimation.

	Gaussian/Student-t		Gumbel/Clayton	
Method	C_P^{Ga}	$C_{\nu,P}^t$	C_θ^{Gu}	C_θ^{Cl}
ρ	0.367	0.397	–	–
ν	–	5.356	–	–
θ	–	–	1.311	0.511
LL	176.631	224.780	177.438	161.183
AIC	-351.261	**-445.560**	-352.875	-320.367

Table 15: 6758.T-9984.T Copula Estimation.

	Gaussian/Student-t		Gumbel/Clayton	
Method	C_P^{Ga}	$C_{\nu,P}^t$	C_θ^{Gu}	C_θ^{Cl}
ρ	0.458	0.466	–	–
ν	–	5.836	–	–
θ	–	–	1.391	0.716
LL	287.481	320.029	257.623	280.106
AIC	-572.963	**-636.059**	-513.246	-558.211

Let us now examine some quantitative measures of dependence. Measures of dependence are common instruments to summarize a complicated dependence structure in a single numerical value. Dependence measures can be categorized into three key concepts. The most well-known is the linear (Pearson) correlation coefficient, defined as

$$\mathsf{Corr}(X_1, X_2) = \frac{\mathsf{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1) \cdot \mathrm{Var}(X_2)}}.$$

Another dependence measure to be considered is rank correlation. The concept of rank correlation focuses on the ranks of given data rather than the data itself. This approach gives rise to important correlation estimators, such as Kendall's tau and Spearman's rho, which we relate to copulas. Because rank-based measures are scale-invariant, they are particularly useful for fitting copulas to empirical data.

Table 16 presents the results of both the linear and rank-based dependence measures for the four pairs of stock returns. All stock pairs show positive dependence across Pearson, Kendall and Spearman measures. The strongest relationships are seen between AAPL–MSFT, indicating a strong linear and monotonic link. Other pairs, like SAP.DE–ADS.DE and 6758.T–9984.T, show moderate dependence, while 0700.HK–0005.HK displays the weakest.

Table 16: Dependence Measures (Pearson, Kendall, Spearman).

Pair	Pearson	Kendall	Spearman
AAPL - MSFT	0.687	0.472	0.637
SAP.DE - ADS.DE	0.436	0.345	0.486
0700.HK - 0005.HK	0.315	0.264	0.376
6758.T - 9984.T	0.444	0.309	0.441

Finally, we introduce the tail dependence of copulas, where we distinguish between upper and lower tail dependence. As observed in the earlier copula plots, there is clear evidence of tail dependence. Specifically, the upper right corner of a copula plot reflects upper tail dependence, indicating co-movement during extreme positive events, while the lower left corner captures lower tail dependence, reflecting joint extreme negative movements.

Tail dependence is formally quantified through tail dependence coefficients, denoted by λ_L and λ_U for the lower and upper tails, respectively. These coefficients measure the probability that one variable takes on an extreme value given that the other does as well. Below, we summarize the theoretical expressions of tail dependence for our used copula families:

- **Gaussian copula**:

$$\lambda_L = \lambda_U = 0 \quad \text{if} |\rho| < 1$$

- **Student-t copula**:

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

  with $t_{\nu+1}(x)$ the CDF of the Student-t distribution. Tail dependence exists if $\rho > -1$.

- **Gumbel copula**:

$$\lambda_U = 2 - 2^{1/\theta}, \quad \lambda_L = 0$$

- **Clayton copula**:

$$\lambda_L = \begin{cases} 2^{-1/\theta} & \text{if } \theta > 0 \\ 0 & \text{else} \end{cases}, \quad \lambda_U = 0$$

The estimated tail dependence coefficients for the four pairs of stock returns are displayed in Tables 17–20. First of all, as expected, the Gaussian copula consistently yields zero tail dependence, confirming its inability to capture joint extreme events. The Student-t copula displays low symmetric tail dependence for most pairs, except for the AAPL–MSFT pair, indicating a strong likelihood of simultaneous extreme returns. Lastly, the Gumbel and Clayton copulas reveal moderate upper and lower tail dependence across all pairs, respectively, with the strongest upper and lower tail dependence both observed for the AAPL–MSFT pair. These quantitative results are consistent with the visual patterns observed in Figures 11–18. In particular, tail dependence is most visually apparent for the AAPL–MSFT pair, which aligns with its relatively high upper and lower tail dependence coefficients.

Table 17: λ_L and λ_U for AAPL-MSFT.

Copula	λ_L	λ_U
Gaussian	0.0000	0.0000
Student-t	0.4282	0.4282
Gumbel	0.0000	0.5454
Clayton	0.6052	0.0000

Table 18: λ_L and λ_U for SAP.DE-ADS.DE.

Copula	λ_L	λ_U
Gaussian	0.0000	0.0000
Student-t	0.2141	0.2141
Gumbel	0.0000	0.3948
Clayton	0.4084	0.0000

Table 19: λ_L and λ_U for 0700.HK-0005.HK.

Copula	λ_L	λ_U
Gaussian	0.0000	0.0000
Student-t	0.1457	0.1457
Gumbel	0.0000	0.3031
Clayton	0.2576	0.0000

Table 20: λ_L and λ_U for 6758.T-9984.T.

Copula	λ_L	λ_U
Gaussian	0.0000	0.0000
Student-t	0.1597	0.1597
Gumbel	0.0000	0.3543
Clayton	0.3796	0.0000

In conclusion, the copula approach does provide significant added value over more standard correlation-based methods. It allows for a more complete and flexible representation of dependencies between financial assets, particularly in capturing asymmetric and tail risk behaviors that are otherwise missed. As such, copulas offer a powerful complement to the standard correlation measures, making them highly valuable in the modeling of financial dependencies.

6 EVT

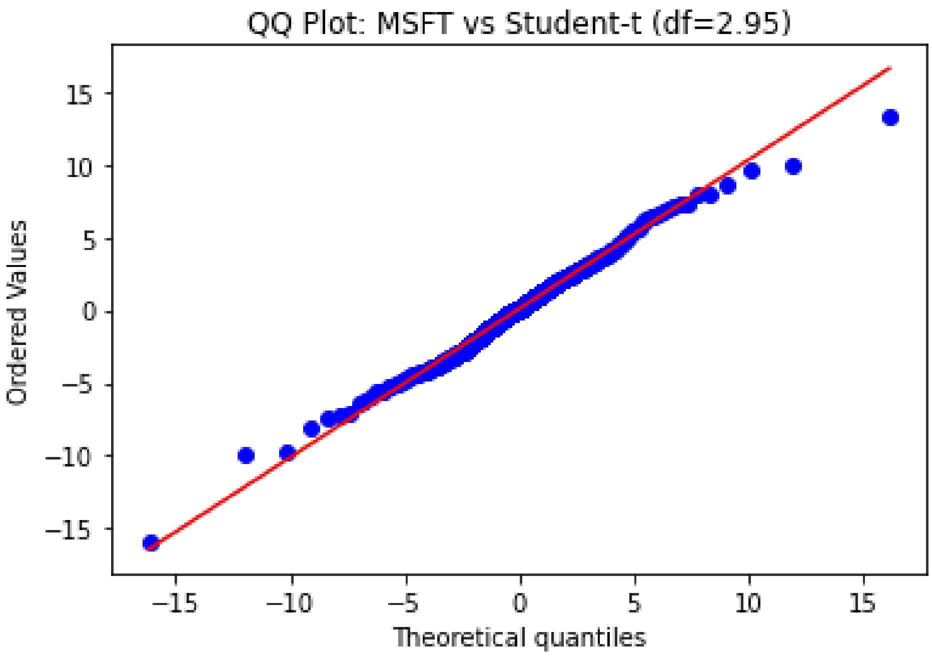


Figure 1: QQ plot for the index 'MSFT'

Figure 19 and Figure 20 located in Appendix A.3 show the QQ plot of 10 asset return series against the Student-t distribution. Our goal is to identify the asset with the heaviest tail, which is crucial for applying Extreme Value Theory (EVT). The heavier the tail, the more the QQ plot deviates from the straight line and forms an S-shape, especially in the tails. This indicates that the asset exhibits more extreme values than would be expected under the reference distribution. Additionally, the degree of freedom parameter in the Student-t distribution provides another indication: the smaller the degree of freedom, the heavier the tail. Among the 10 assets, the MSFT index has the lowest degree of freedom, estimated at 2.95, and its QQ plot 1 shows the most pronounced S-curve. Based on these two observations, we select MSFT as the asset with the heaviest tail.

Before we utilize EVT, it is important to note that we work on the loss side, meaning we take the negative of the returns. This transformation ensures that losses appear as positive values. Accordingly, all exceedances and related analyses are also conducted on this loss basis. Specifically, the exceedance over a threshold u is defined as:

$$X_u = X - u \quad \text{given that} \quad X > u$$

where X represents the loss.

To apply EVT effectively, choosing an appropriate threshold is crucial. This threshold determines which

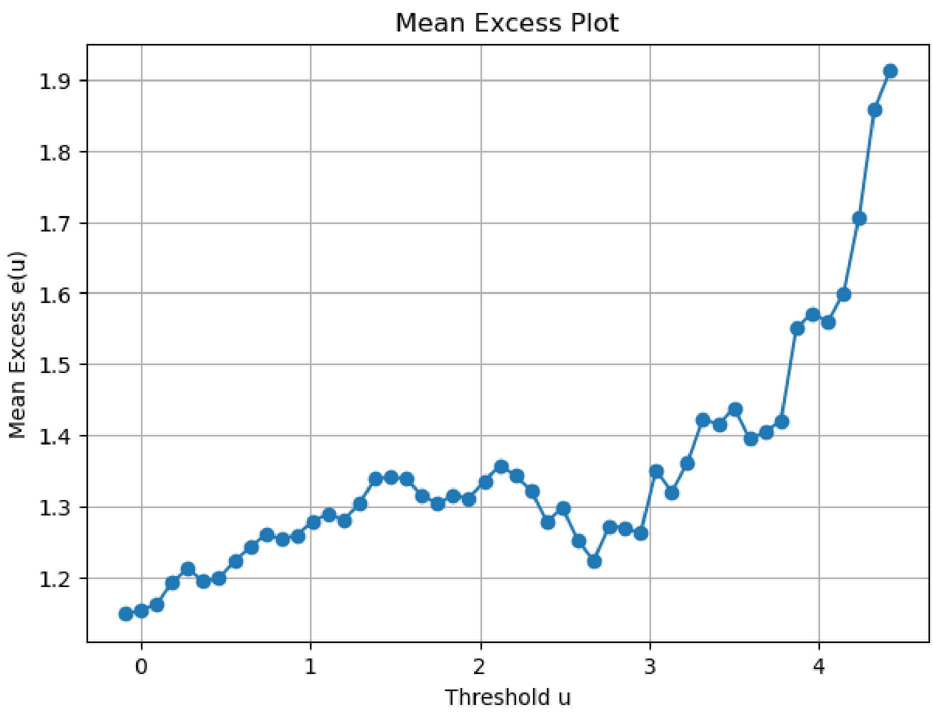


Figure 2: Mean excess function versus threshold for MSFT returns. A linear trend beyond suggests this as a suitable threshold.

data points are considered extreme and thus modeled using the Generalized Pareto Distribution (GPD). If the threshold is too low, the GPD assumption may not hold. If it's too high, the sample size becomes too small. We choose our threshold based on the mean excess plot shown in Figure 2. According to EVT theory, if the GPD is a good model for the exceedances, the mean excess function should appear approximately linear with a positive slope beyond a certain threshold. This is because, for a GPD, the mean excess function is linear in the threshold u as following :

$$e(u) = \frac{\beta(u)}{1-\xi} = a + \frac{\xi}{1-\xi} \cdot u \tag{1}$$

While the plot does not show a perfect straight line, it becomes approximately linear with a positive slope after a threshold of 2.6 and we chose 2.6 to be our threshold. To validate this choice, we fit the GPD to the exceedances over the threshold 2.6. The resulting fit in Figure 3 appears reasonable, suggesting that this threshold is appropriate. The estimated GPD parameters are $\xi = 0.1999$ and $\beta = 0.9885$, which indicates a heavy-tailed distribution (since $(\xi > 0)$.).

With these parameter estimates, we compute Value-at-Risk (VaR) and Expected Shortfall (ES) using EVT.

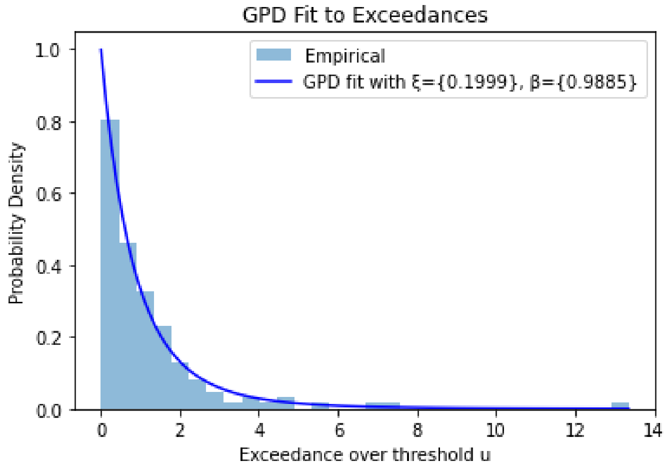


Figure 3: Histogram of exceedances above $u = 2.6$ (bars) with the fitted GPD density (solid line).

The formulas used are defined as follows:

$$\mathrm{VaR}_\alpha = u + \frac{\beta}{\xi}\left(\left(\frac{1-\alpha}{\overline{F}(u)}\right)^{-\xi} - 1\right), \tag{2}$$

$$\mathrm{ES}_\alpha = \frac{\mathrm{VaR}_\alpha + \beta - \xi u}{1-\xi}, \tag{3}$$

where $\overline{F}(u)$ is the exceedance probability at threshold u. Following Smith Smith, 1987, we estimate it empirically as

$$\widehat{\overline{F}}(u) \approx \frac{N_u}{n},$$

where N_u is the number of exceedances above the threshold u, and n is the total number of observations. In our case, this yields an empirical exceedance probability of approximately $\widehat{\overline{F}}(u) \approx 0.0555$.

Table 21: VaR and ES Comparison by Method (Loss Basis)

Method	**VaR (95%)**	**ES (95%)**	**VaR (99%)**	**ES (99%)**
EVT (GPD)	2.7052	3.9670	4.6218	6.3625
Historical Simulation	2.7182	3.9637	4.4469	6.3329
GARCH (Student-t)	2.2763	3.4862	4.1019	5.7629

For comparison, we also compute VaR and ES using Historical Simulation and GARCH with Student-t innovations. Since EVT is specifically designed to capture tail risk, we expect it to yield more conservative risk estimates, especially for a heavy-tailed asset like MSFT. As shown in Table 21, EVT produces the highest VaR and ES values among the three methods, confirming that it captures the potential for extreme losses more effectively. Notably, the results from Historical Simulation are quite close to those of EVT, particularly at the 99% level, suggesting that the empirical distribution of MSFT already reflects a significant degree of tail heaviness.

7 Recommendations for Further Actions

First, the part about PCA is discussed. In general PCA is a relatively robust method. As indicated by the math in Appendix A.1, it simply does what it is supposed to do and is a deterministic derivation of the principal components from where it is rather straight forward to select the PC's which explain a reasonable percentage of the variance in the data. One variation on the standard approach of PCA that one can consider is to make segments of variables before applying PCA to then from these select the PC's that explain most variance in these segments of variables. Nevertheless, the orthogonality of the PC's then no longer holds so it is unlikely that this would lead to an average higher variance explained per PC included. This being said, it would make it easier to classify PC's on what variables within the dataset they represent as the factor loading analysis to dissect this from the PC's is quite a hand wavy method, so there it would improve the clarity on what variables each PC does represent. This is a trade-off one can make depending on the goal of the research, in the context of this specific study where we investigate the relationships between several assets perhaps this segmenting strategy is more appropriate as it will make it easier to identify certain relationships within the data. Although it does make it more difficult to identify specific shared relationships in the data. Therefore a careful consideration is needed to make this decision appropriately.

Second, the factor analysis is discussed. In this report, we conduct two dimensions of factor analysis, macro factor method and statistical method based on the PCA approach. However, for the macro factor approach, the constructed regression results show the weakness coming from current factor selection, especially for country-wise characteristics, which brings the reflection that more solid macro indicators can be included, such as the more accurate regional volatility proxy, specified regional Fama-French factors and the alternatives to the currency exchange rate. For the statistical method, although the statistical results outperform the former, the explanatory power is limited and the country and industry specific factors should be also included to boost the analysis and provide the implications for global diverse portfolio management, also the correlations between factors can bring impact such as multicollinearity towards the model accuracy.

Third, the copula part is discussed. The copula approach offers substantial advantages over traditional correlation-based methods. It enables a more comprehensive and flexible modeling of dependencies between financial assets, particularly in capturing asymmetric relationships and tail risk which is often overlooked by standard techniques. As a result, copulas serve as a powerful complement to the standard correlation measures, enhancing the accuracy and robustness of financial dependency modeling.

Lastly, the EVT part was discussed. In applying EVT, we naturally assumed that the data is independently and identically distributed (i.i.d.) to satisfy the theoretical requirements. However, whether financial returns are truly i.i.d. remains questionable, as they typically exhibit time dependence and volatility clustering. It is important to note that Value-at-Risk (VaR) and Expected Shortfall (ES) can be obtained through a location and scale transformation. This implies that even if the original return data is not i.i.d., we can standardize it to approximate i.i.d. behavior and then apply EVT. For example, one can estimate an ARMA-GARCH model and extract standardized residuals using

$$Z_t = \frac{X_t - \mu_t}{\sigma_t}.$$

EVT (GEV or GPD) can then be applied to the residuals to compute VaR or ES. Finally, the risk measures can be transformed back to the original scale using the inverse of the location and scale transformation. In this project, EVT was applied directly to raw returns. For future improvement, applying EVT to standardized residuals obtained from a GARCH-type model could enhance the robustness and theoretical validity of the tail risk estimates.

References

French, K. R. (2024). Fama/french developed markets factors [daily] [Accessed: 2024-05-13]. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Smith, R. L. (1987). Estimating tails of probability distributions. *Annals of Statistics*, *15*(3), 1174–1207.

A Appendix

A.1 Principal Component Analysis Derivations

Let $Y \in \mathbb{R}^{N \times T}$ be the data matrix of demeaned asset returns where N is the number of assets and T is the number of time periods. Each column of Y represents a time observation and each row corresponds to an individual asset's return over time. The sample covariance matrix $S \in \mathbb{R}^{N \times N}$ is defined as:

$$S = \frac{1}{T-1} Y Y^\top.$$

This matrix captures the empirical second moment structure of the asset returns. To perform PCA we diagonalize the covariance matrix via an eigenvalue decomposition:

$$S = X \Lambda X^\top,$$

where $X \in \mathbb{R}^{N \times N}$ is the orthonormal matrix of eigenvectors and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_N)$ is the diagonal of eigenvalues ordered such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0$. The eigenvectors define the directions of maximum variance and the eigenvalues indicate the magnitude of variance explained along these directions.

The principal components, or factor scores, are obtained by projecting the data onto the eigenvectors:

$$C = X^\top Y,$$

where $C \in \mathbb{R}^{N \times T}$ contains the time series of the principal components. The variance of the components is given by:

$$\frac{1}{T-1} C C^\top = X^\top \left(\frac{1}{T-1} Y Y^\top \right) X = X^\top S X = \Lambda,$$

which confirms that the variance of each principal components equals its corresponding eigenvalue. This result also implies that the principal components are uncorrelated as X is an orthonormal matrix as defined above. This is used in the equation above to know that $X^\top S X = X^\top X S X^\top X = I_N \Lambda I_N = \Lambda$ because one of the properties of an orthonormal matrix is that the inner product of it equals the identity matrix indicated by I_N here.

To approximate the original data using the first $k < N$ components we define $X_k \in \mathbb{R}^{N \times k}$ as the matrix of the first k eigenvectors and compute the rank-k approximation of the data as:

$$\widetilde{Y} = X_k X_k^\top Y.$$

This projection minimizes the reconstruction error in the least squares sense among all rank-k approximations. Therefore, PCA can be viewed as a spectral decomposition of the sample covariance matrix leading to a possibility for dimension reduction of the data while maintaining maximum variance of the data. The principal directions (eigenvectors) and corresponding variances (eigenvalues) arise directly from the second moment structure of the data.

A.2 Additional results

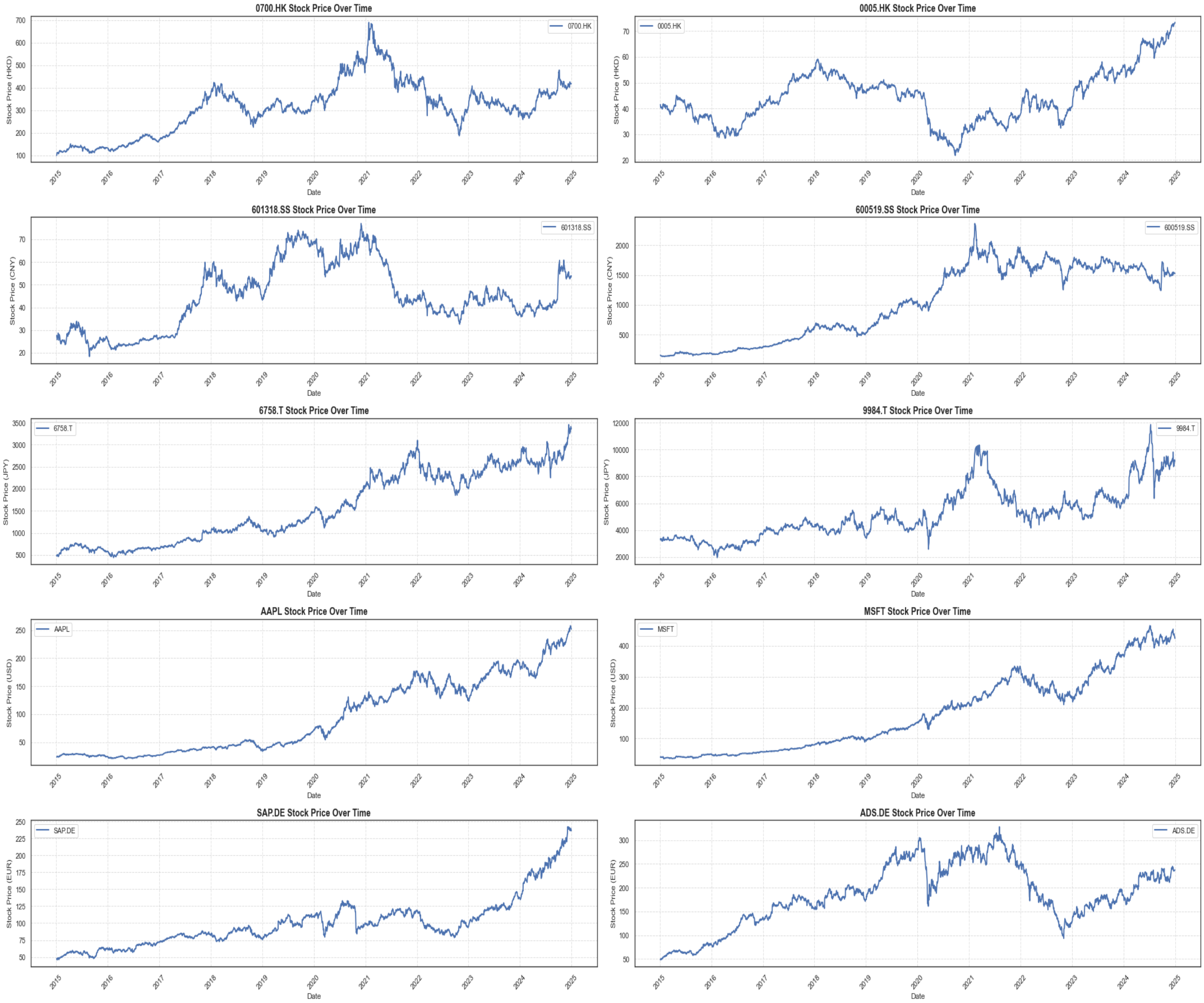


Figure 4: Stock price evolution from January 1, 2015, to December 31, 2024, for the 10 assets: Apple (AAPL, USD), Microsoft (MSFT, USD), SAP (SAP.DE, EUR), Adidas (ADS.DE, EUR), Sony (6758.T, JPY), SoftBank Group (9984.T, JPY), Tencent (0700.HK, HKD), HSBC Holdings (0005.HK, HKD), Ping An Insurance (601318.SS, CNY) and Kweichow Moutai (600519.SS, CNY).

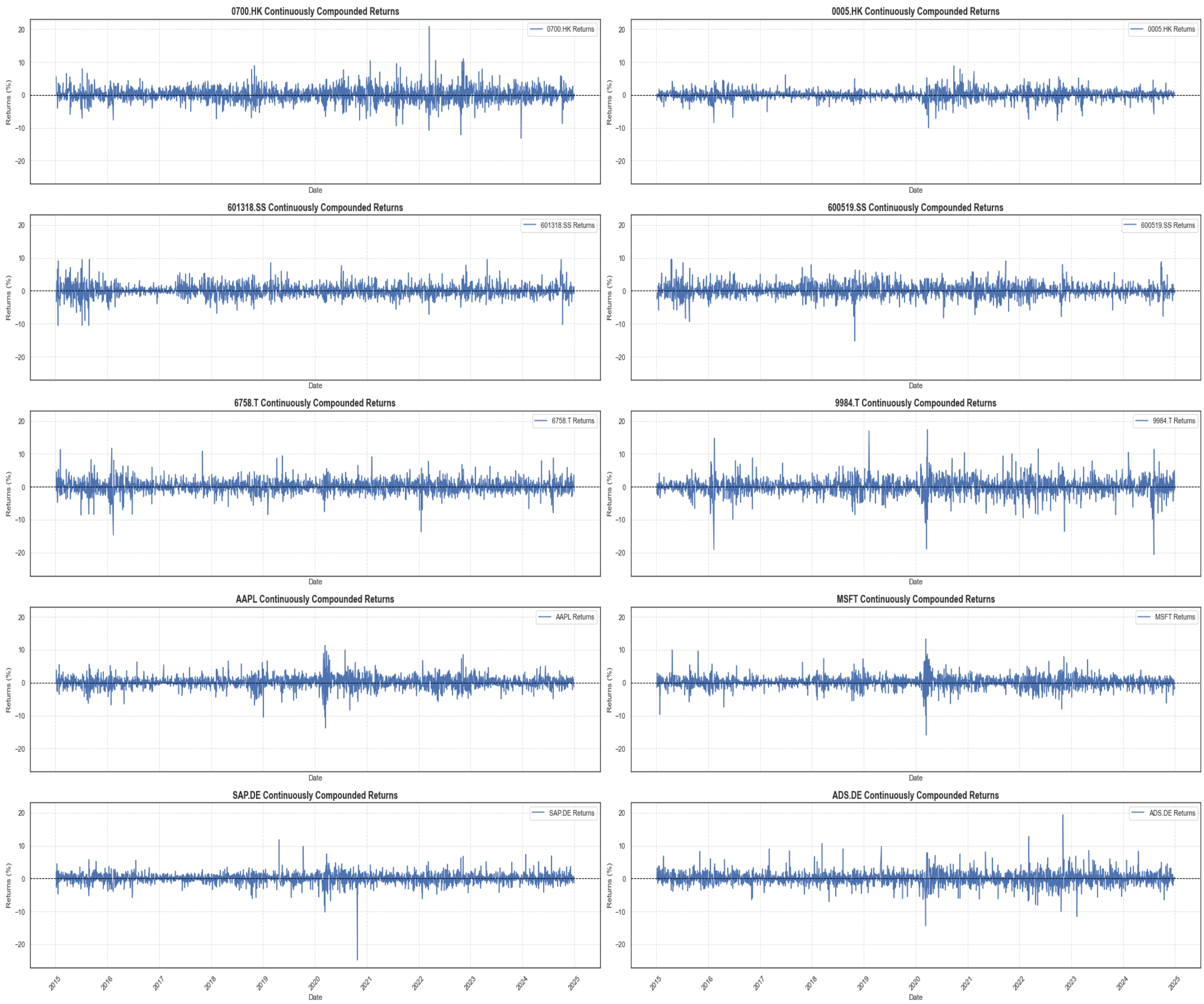


Figure 5: Time series plots of continuously compounded daily returns for the 10 asset constituents, grouped by currency. Each subplot displays the return dynamics of an individual stock from January 1, 2015 to December 31, 2024. Horizontal dashed lines mark the zero-return level. The y-axis scale is unified across all assets to allow direct comparison of return volatility.

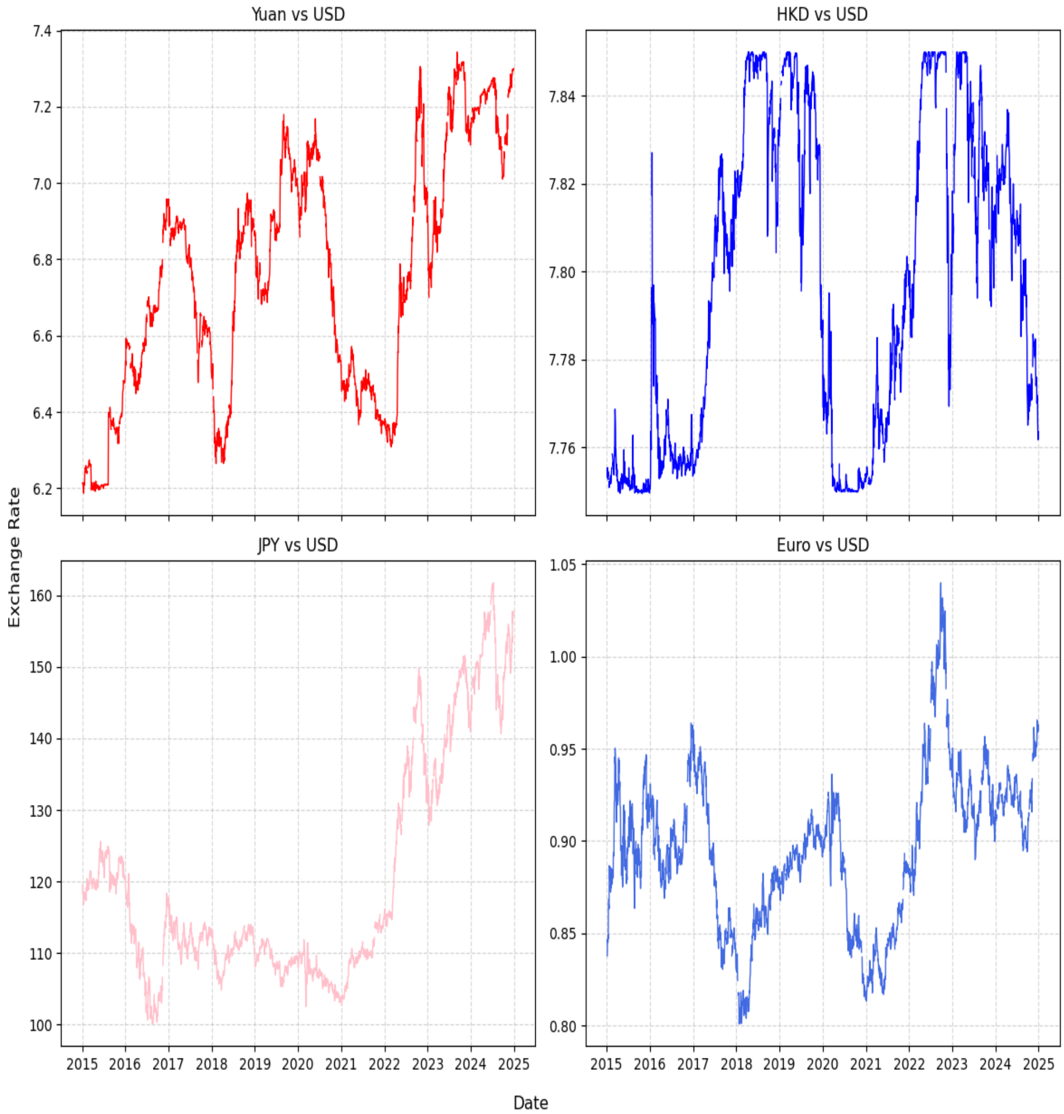


Figure 6: Time series plots of currencies' exchange rate of towards United State Dollars (USD). Including Exchange Rate for People's Republic of China (Yuan/US$), Exchange Rate for Hong Kong (Dollar/US$), Exchange Rate for Japan (Yen/US$), Exchange Rate for European Monetary Union (Euro/US$), from January 1, 2015 to December 31, 2024.

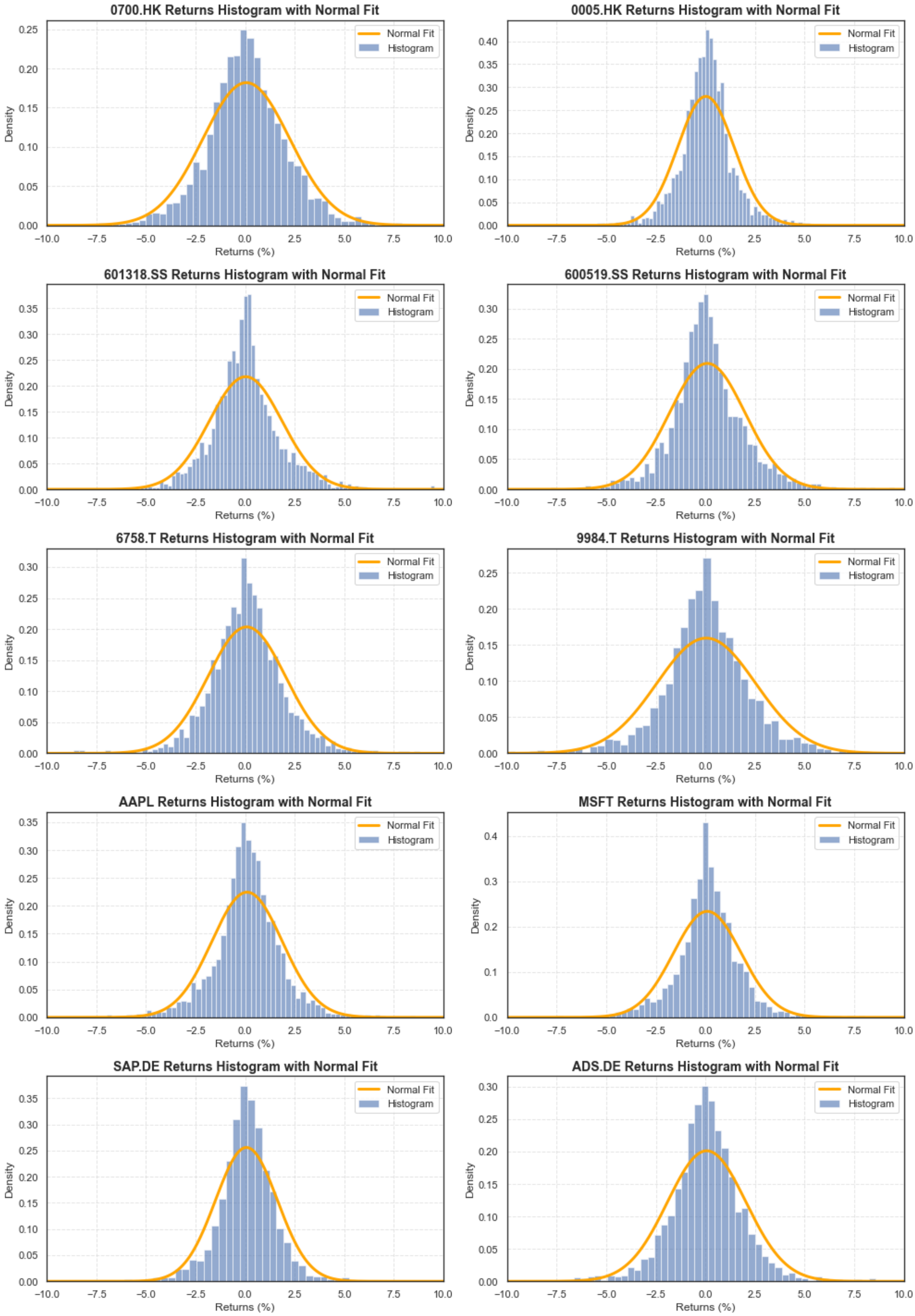


Figure 7: Histograms of daily log returns for the 10 asset constituents with fitted Gaussian distributions. Grouped by currency, each subplot shows the marginal return distribution for one asset along with the Gaussian fit. Sample period: January 1, 2015 to December 31, 2024.

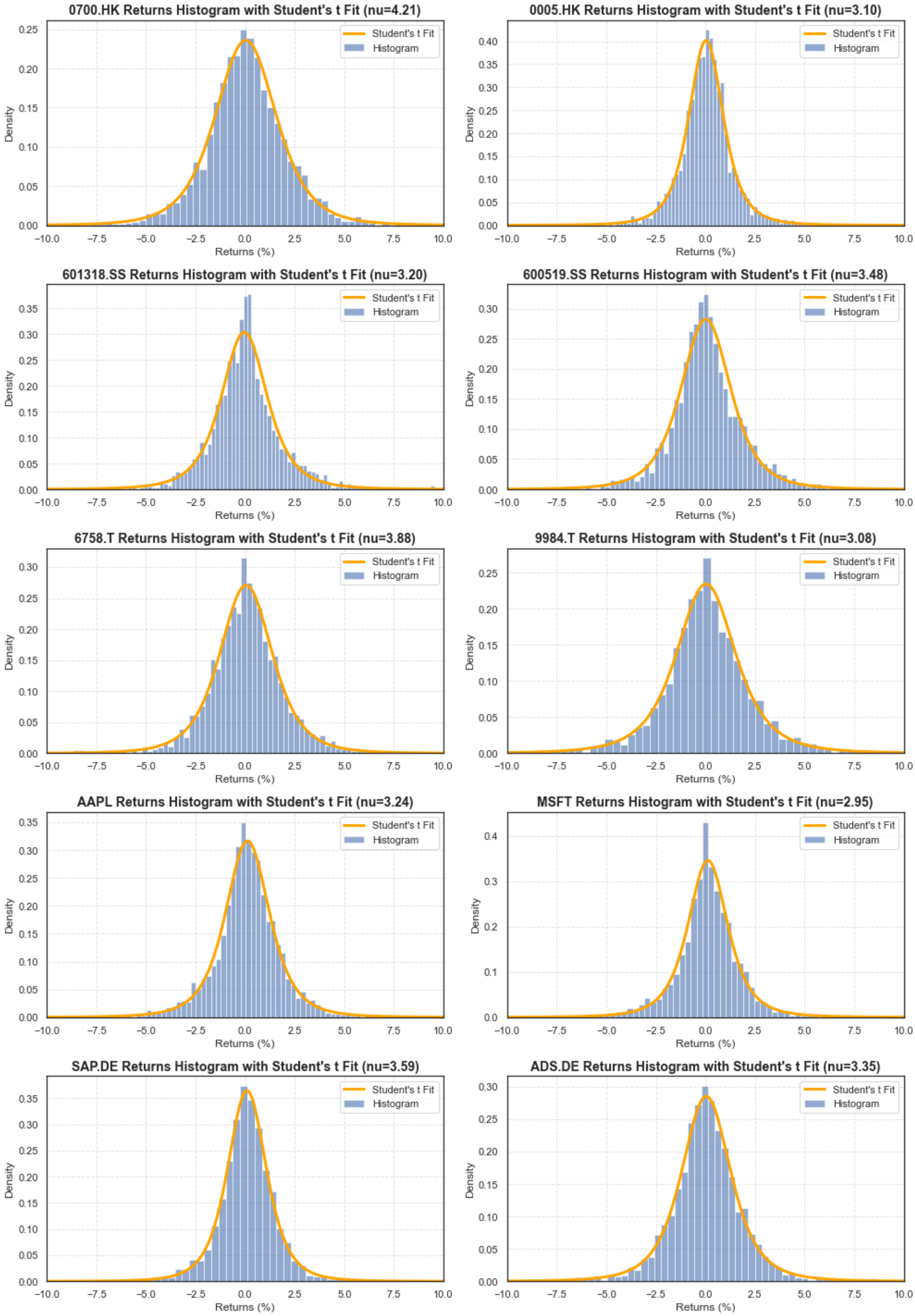


Figure 8: Histograms of daily log returns for the 10 asset constituents with fitted Student's t-distributions. Grouped by currency, each subplot shows the marginal return distribution for one asset along with the heavy-tailed t-distribution fit. The degrees of freedom parameter (ν) is estimated from the data, allowing for non-integer values. Sample period: January 1, 2015 to December 31, 2024.

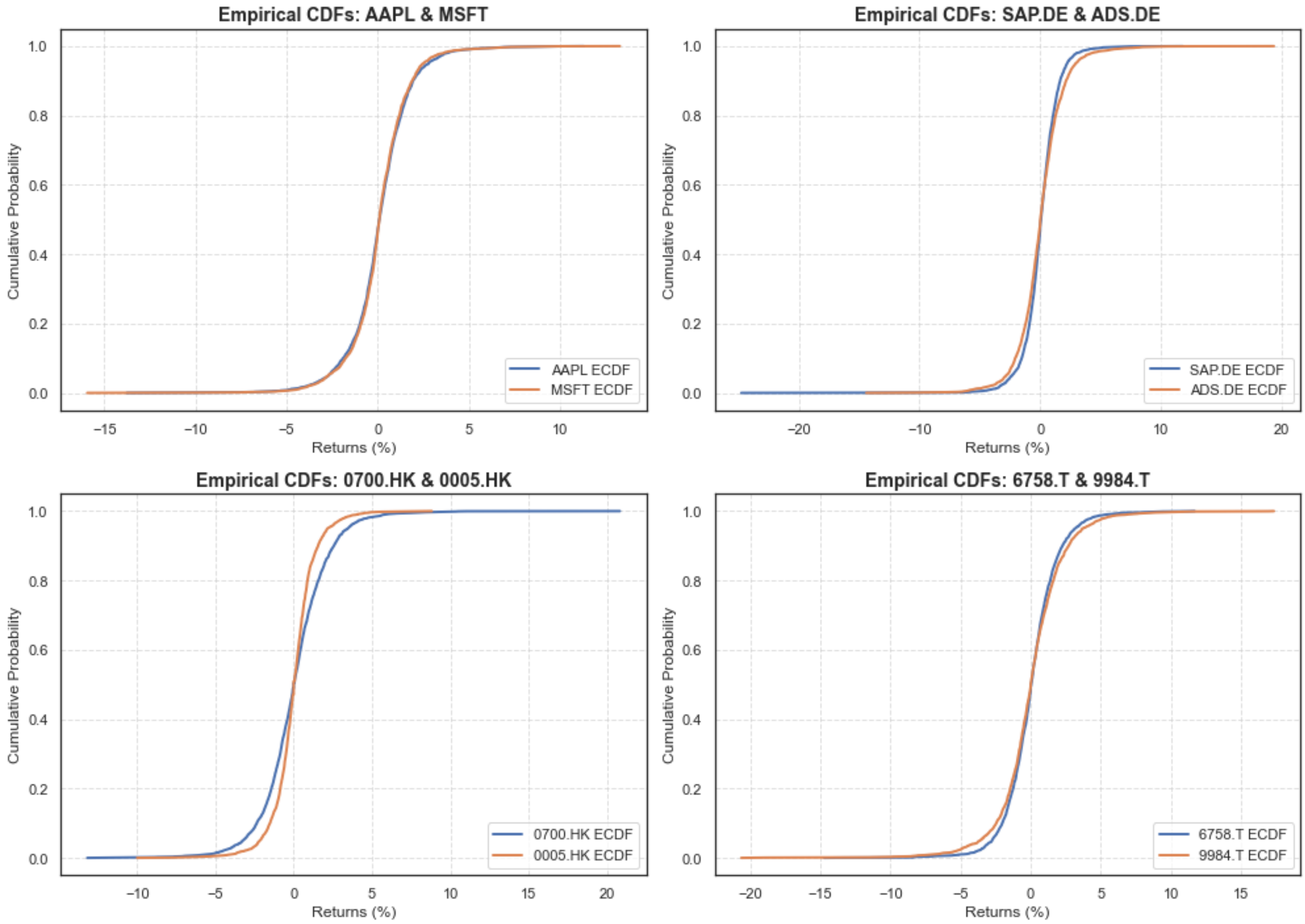


Figure 9: Empirical Cumulative Distribution Functions (ECDFs) for return pairs. Each subplot displays the ECDFs of two assets' daily returns.

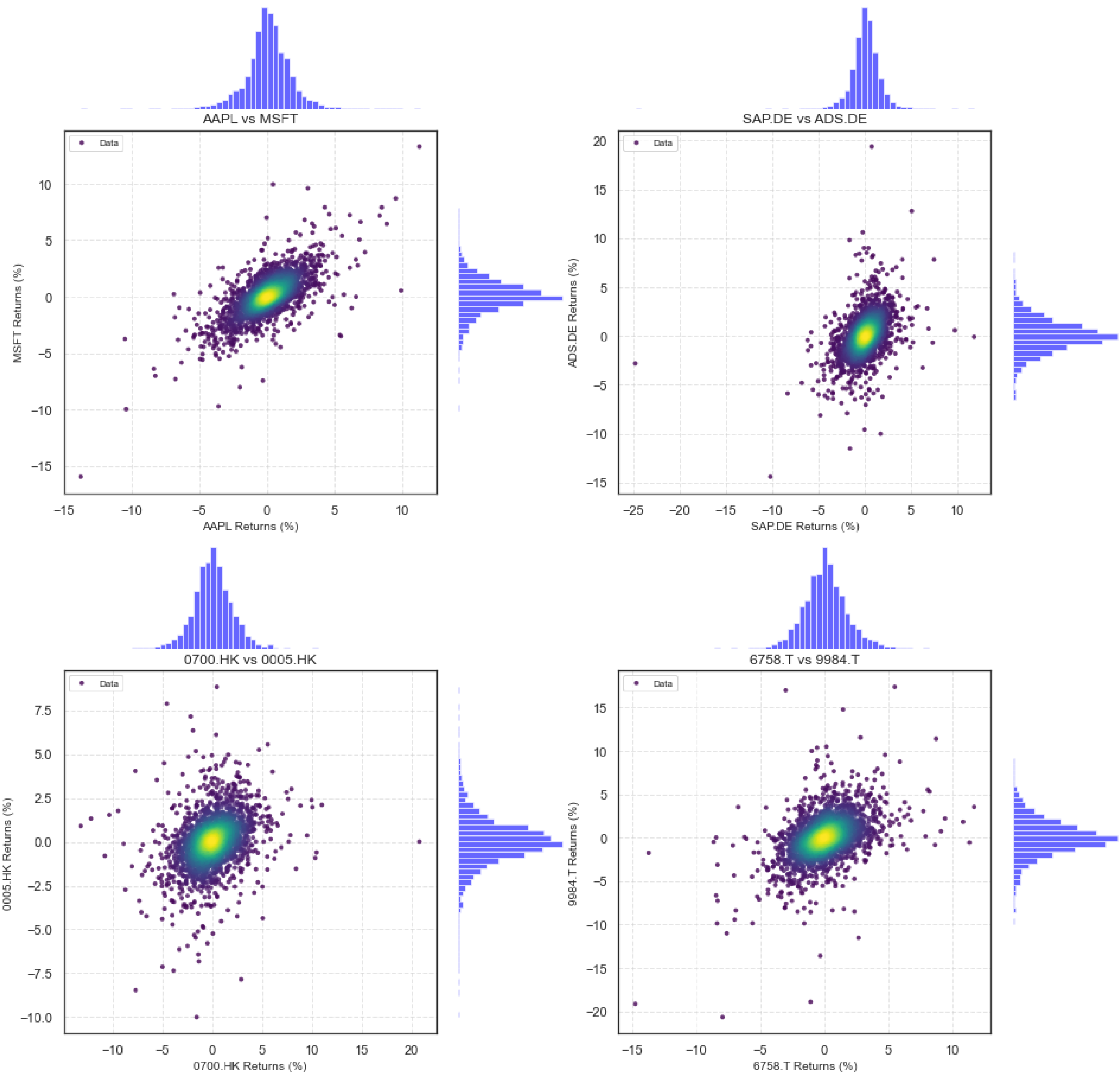


Figure 10: Bivariate scatter plots of daily log returns for selected asset pairs, each complemented with marginal histograms and colored by kernel density estimation (KDE). The assets are grouped by currency and each subplot visualizes the joint return distribution for one pair.

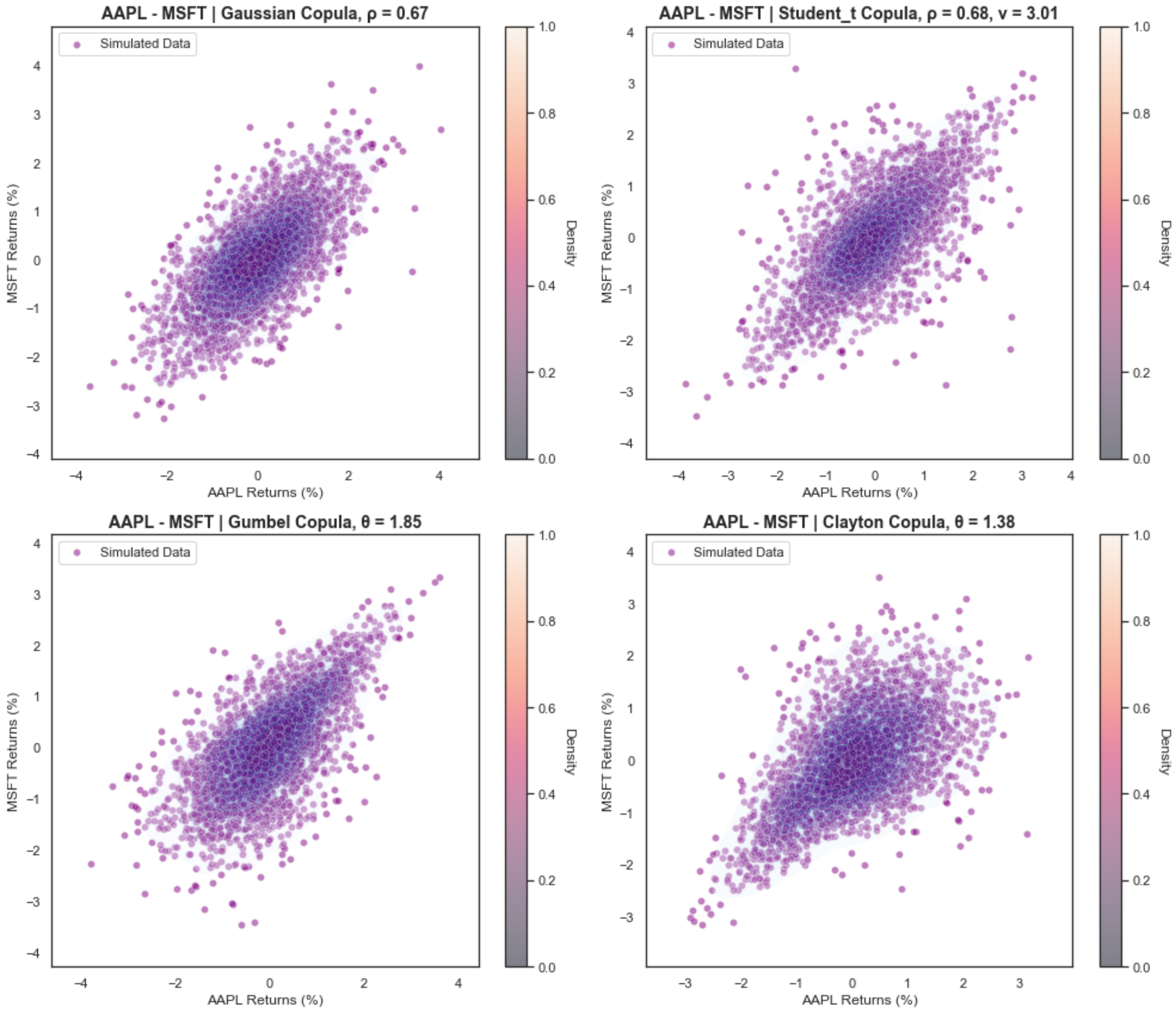


Figure 11: Simulated copula dependence structures for AAPL-MSFT. Each panel shows the scatter and density contours for a different copula family (Gaussian, Student-t, Gumbel and Clayton), providing insight into their ability to model tail and overall dependence.

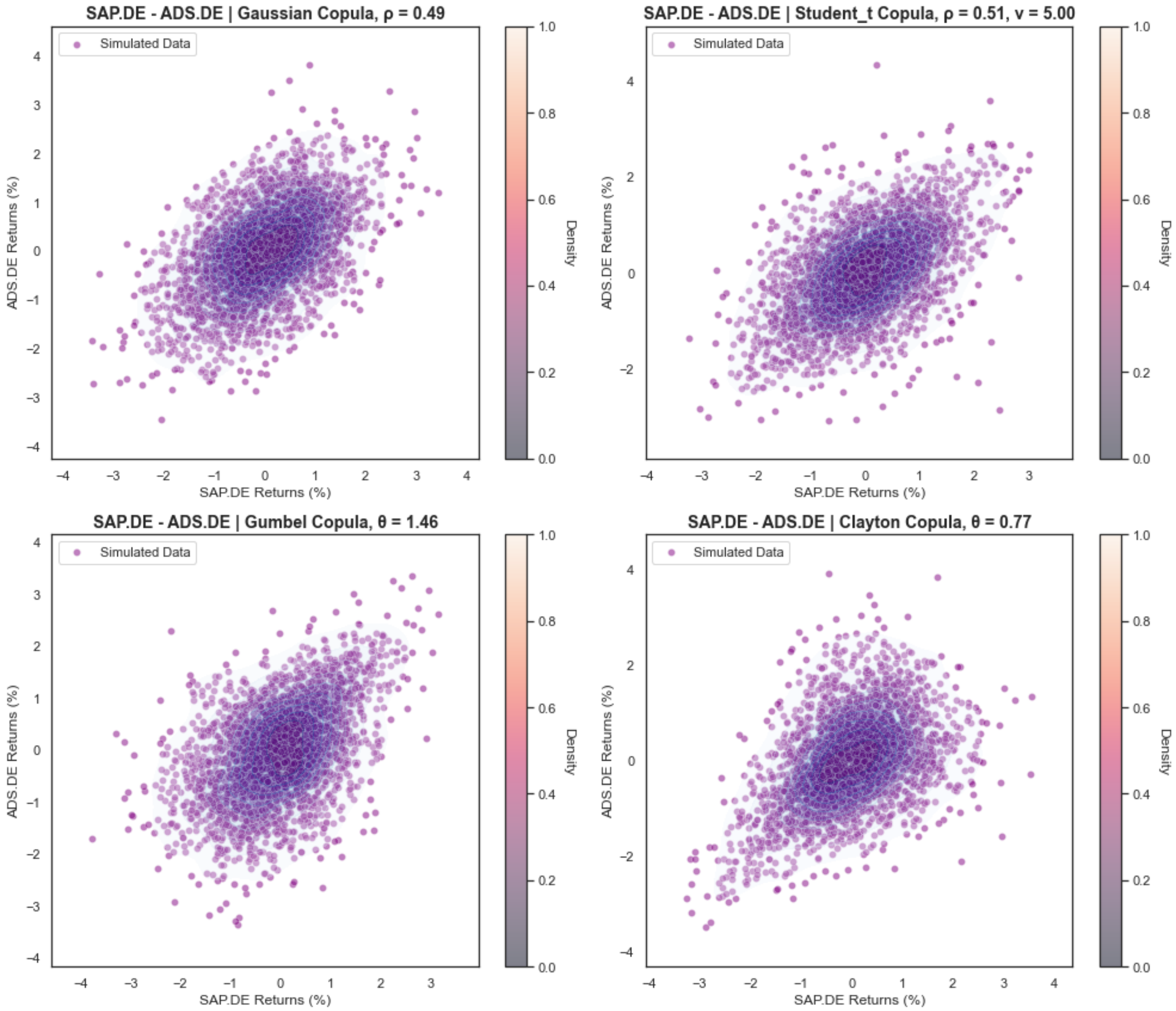


Figure 12: Simulated copula dependence structures for SAP.DE-ADS.DE. Each panel shows the scatter and density contours for a different copula family (Gaussian, Student-t, Gumbel and Clayton), providing insight into their ability to model tail and overall dependence.

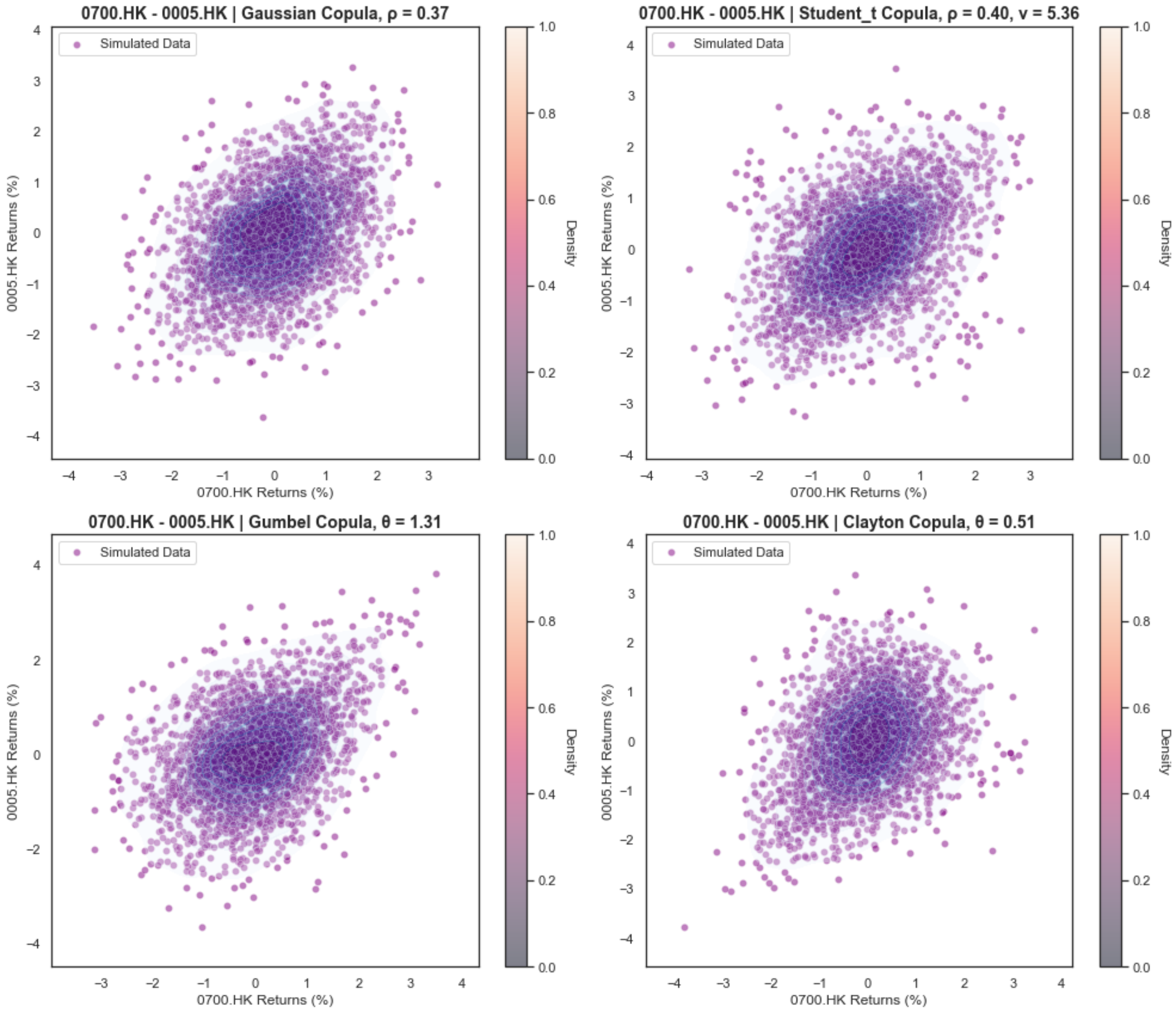


Figure 13: Simulated copula dependence structures for 0700.HK-0005.HK. Each panel shows the scatter and density contours for a different copula family (Gaussian, Student-t, Gumbel and Clayton), providing insight into their ability to model tail and overall dependence.

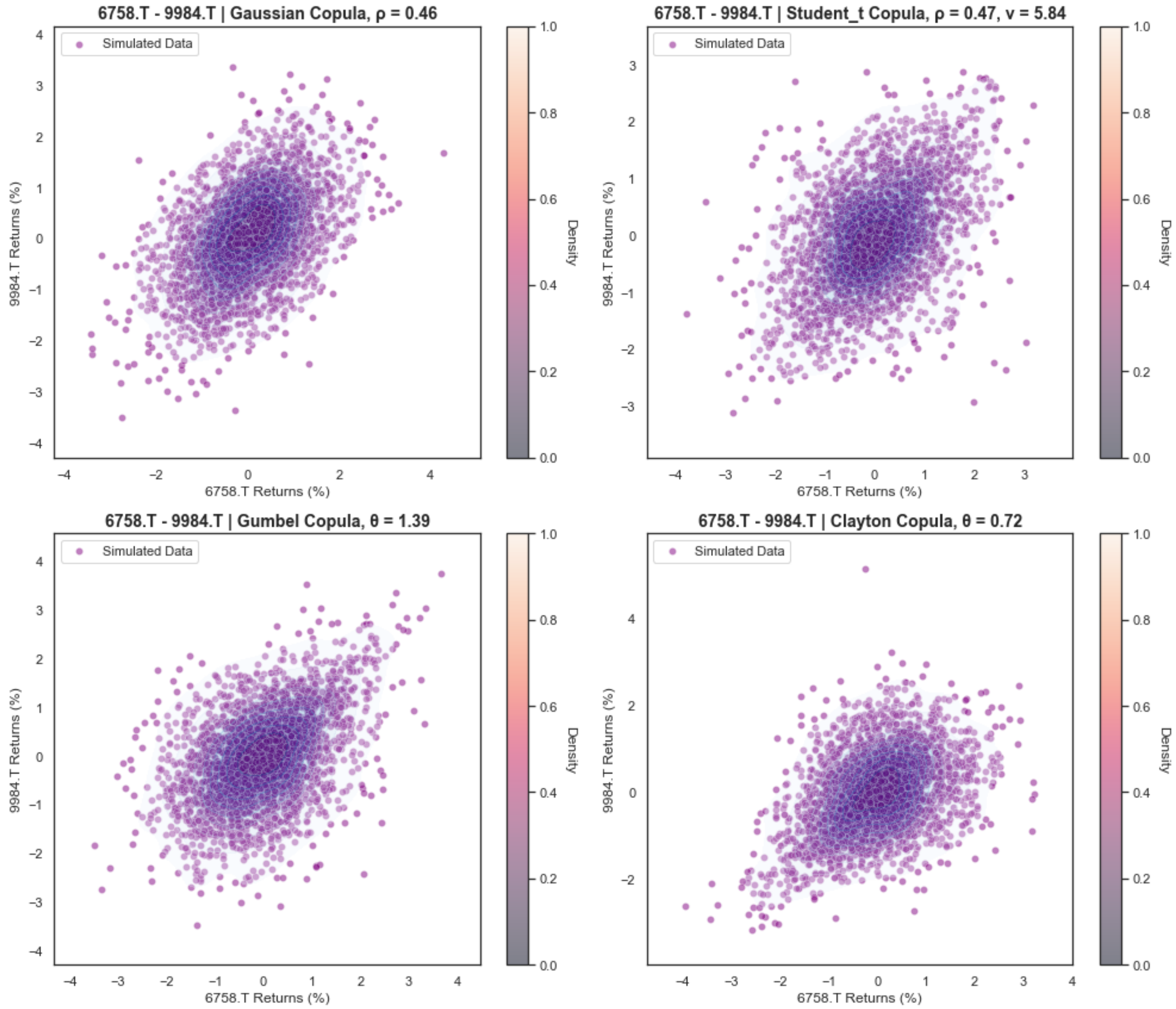


Figure 14: Simulated copula dependence structures for 6758.T-9984.T. Each panel shows the scatter and density contours for a different copula family (Gaussian, Student-t, Gumbel and Clayton), providing insight into their ability to model tail and overall dependence.

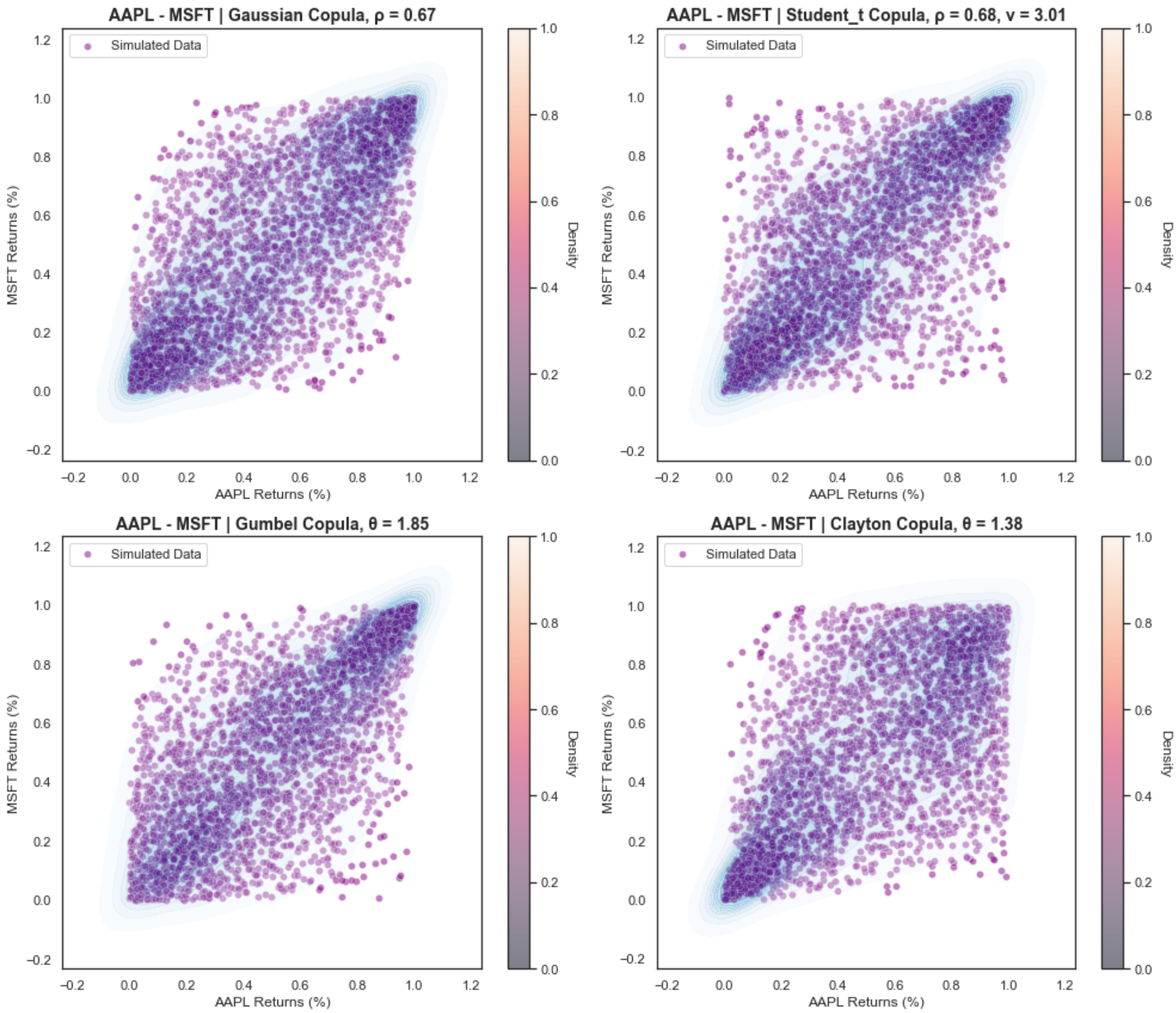


Figure 15: Simulated copula samples on the *uniform scale* for AAPL-MSFT. Each subplot shows a different copula family (Gaussian, Student-t, Gumbel and Clayton) applied to uniform marginals, highlighting dependence structure without marginal transformation.

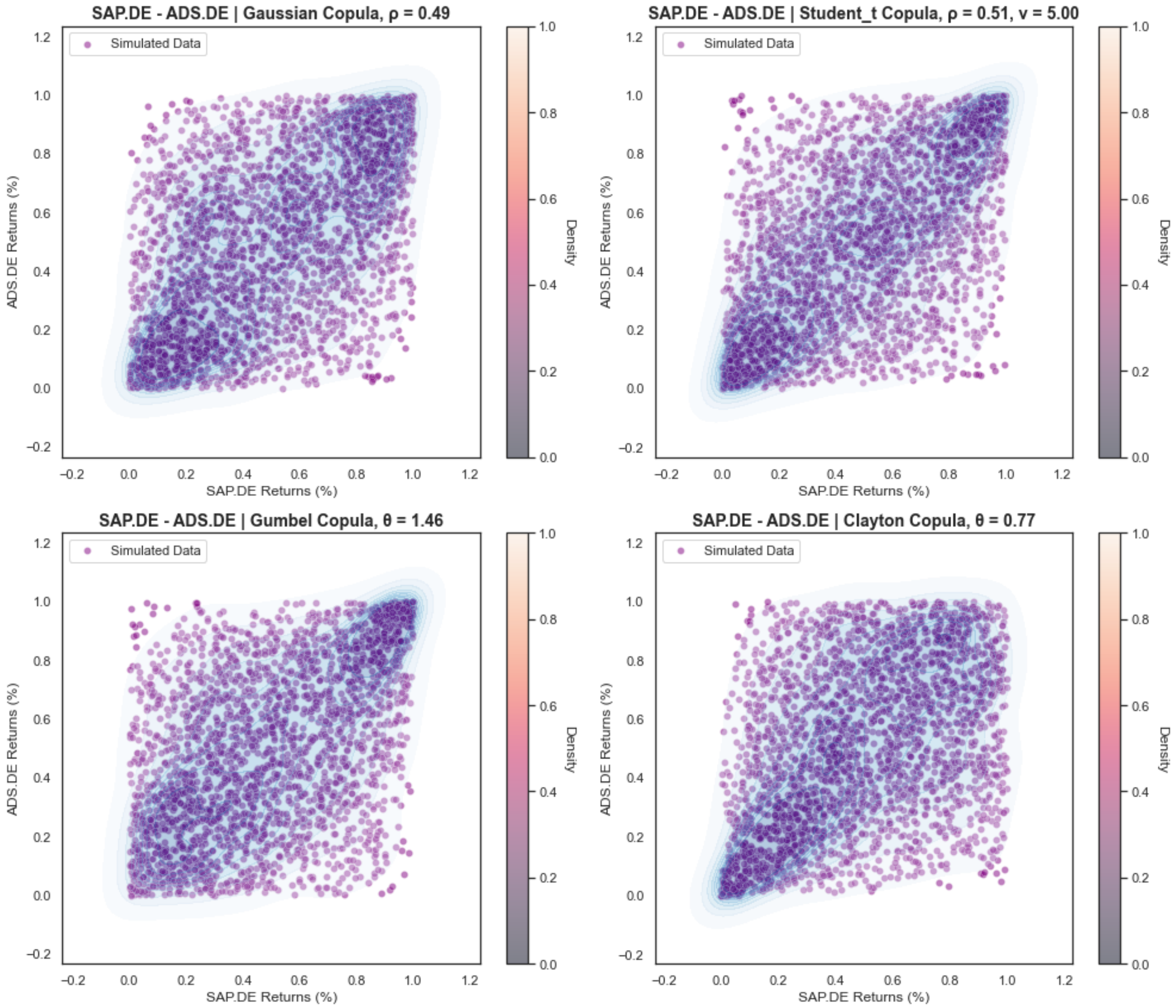


Figure 16: Simulated copula samples on the *uniform scale* for SAP.DE-ADS.DE. Each subplot shows a different copula family (Gaussian, Student-t, Gumbel and Clayton) applied to uniform marginals, highlighting dependence structure without marginal transformation.

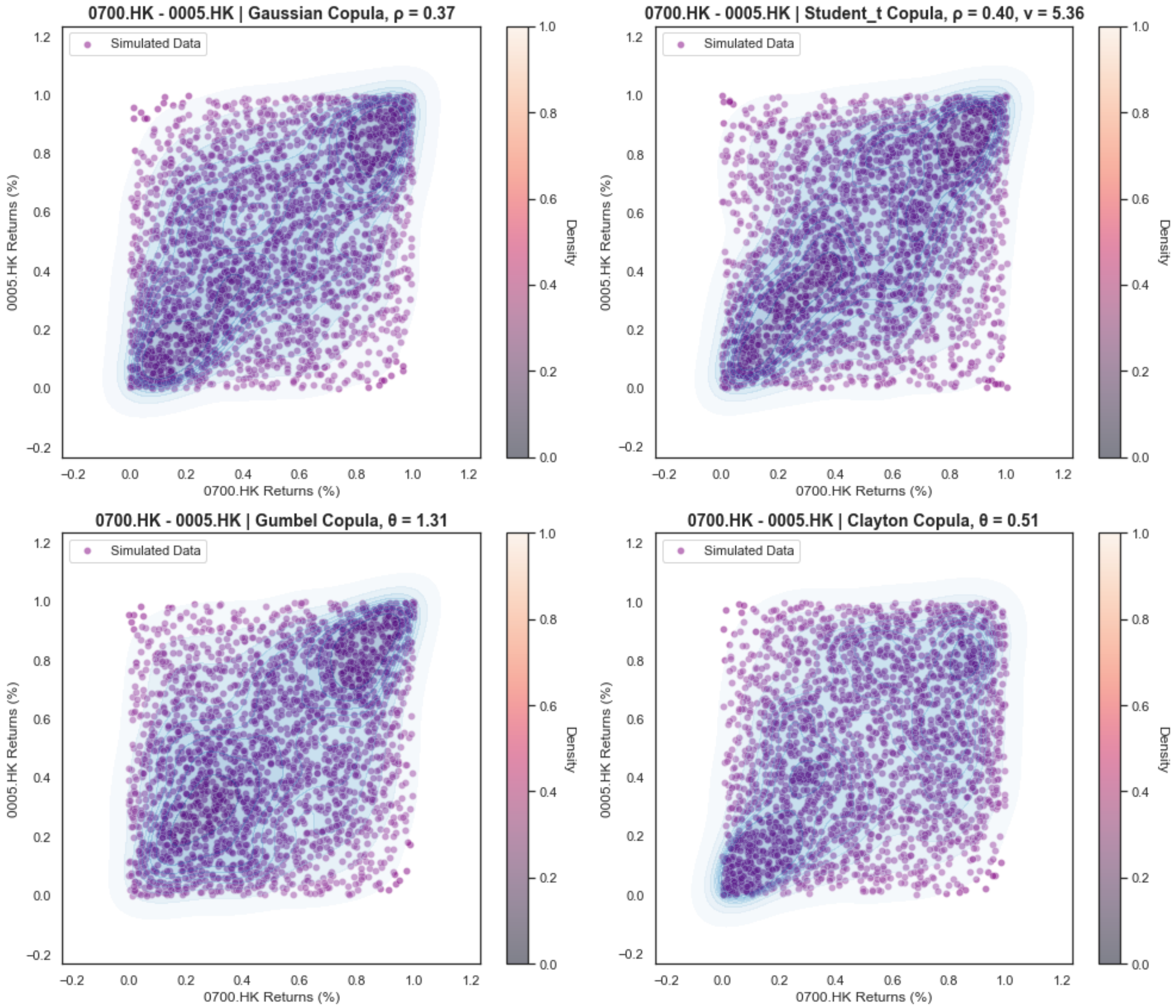


Figure 17: Simulated copula samples on the *uniform scale* for 0700.HK-0005.HK. Each subplot shows a different copula family (Gaussian, Student-t, Gumbel and Clayton) applied to uniform marginals, highlighting dependence structure without marginal transformation.

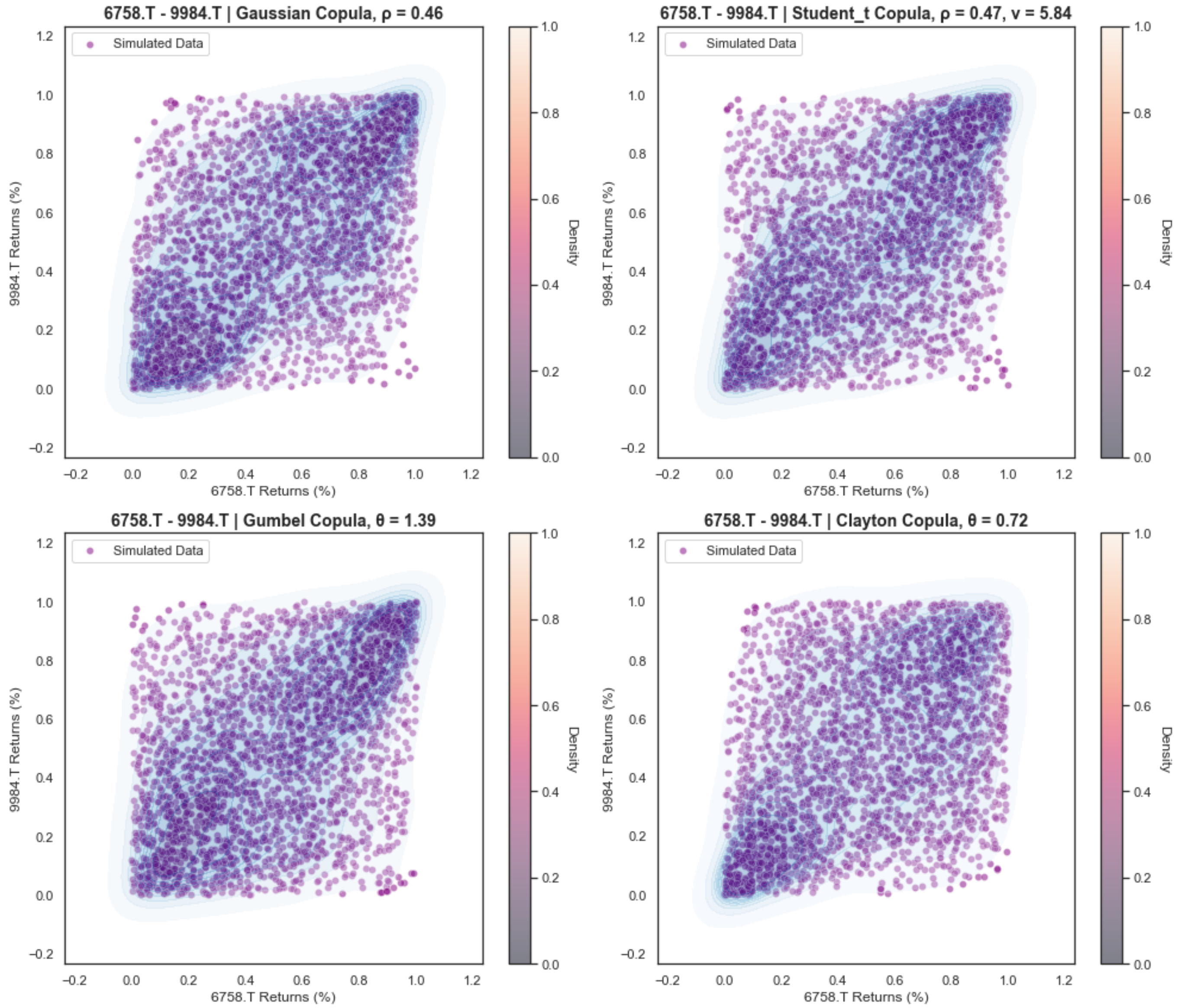
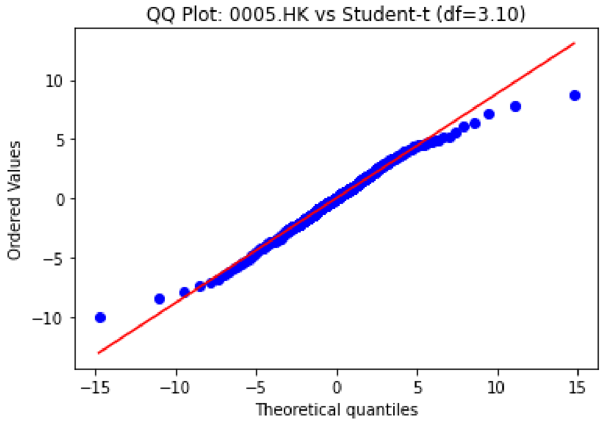
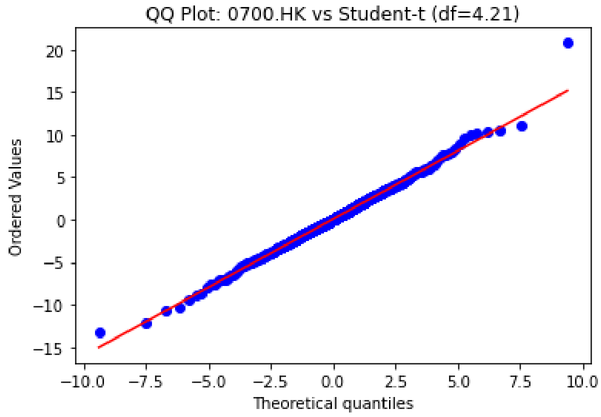


Figure 18: Simulated copula samples on the *uniform scale* for 6758.T-9984.T. Each subplot shows a different copula family (Gaussian, Student-t, Gumbel and Clayton) applied to uniform marginals, highlighting dependence structure without marginal transformation.

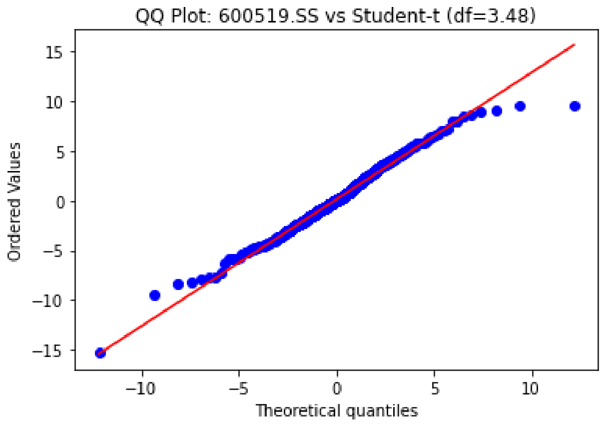
A.3 EVT plots



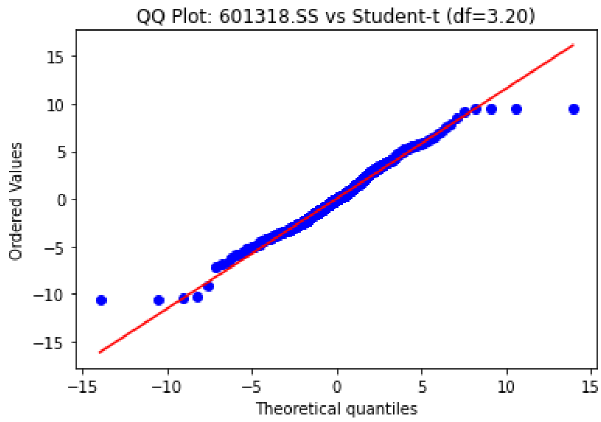
(a) 0005.HK



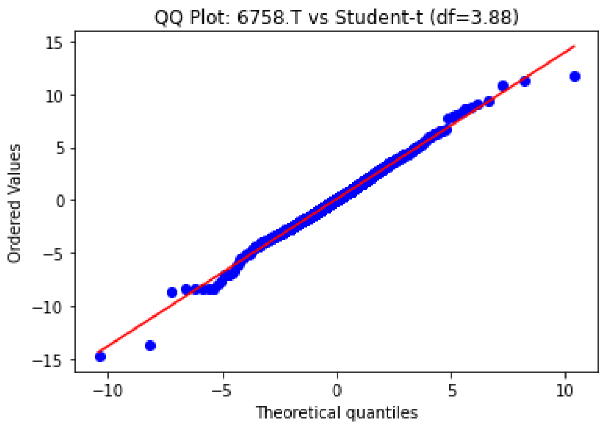
(b) 0700.HK



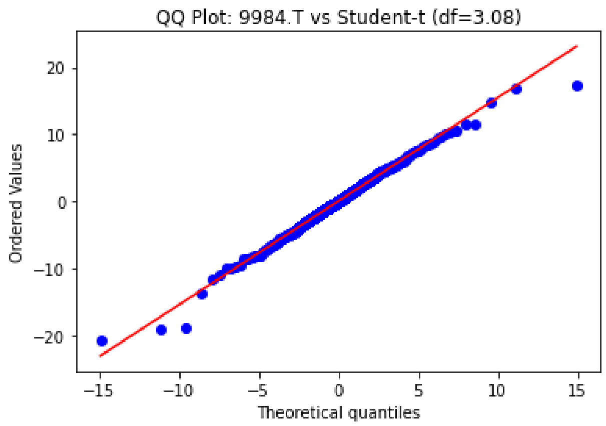
(c) 600519.SS



(d) 601318.SS

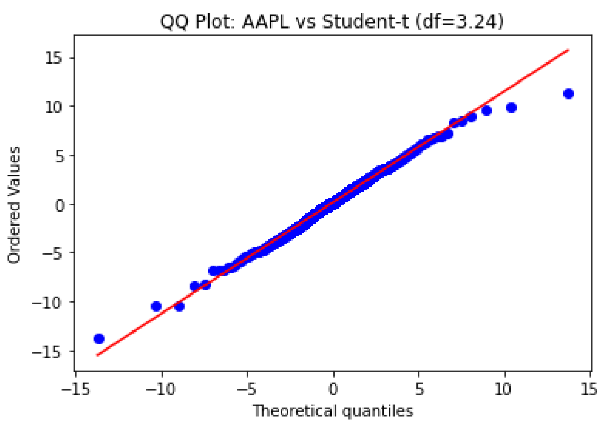


(e) 6758.T

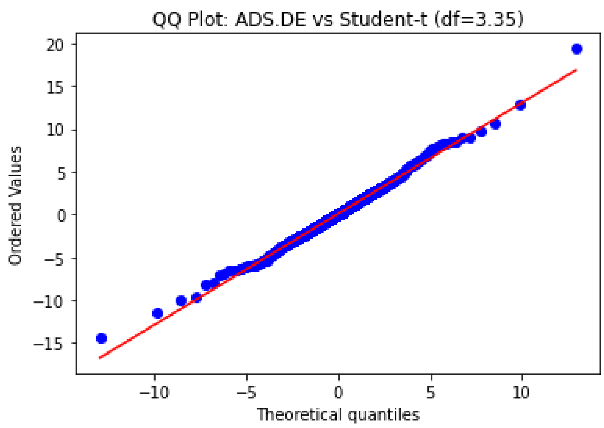


(f) 9984.T

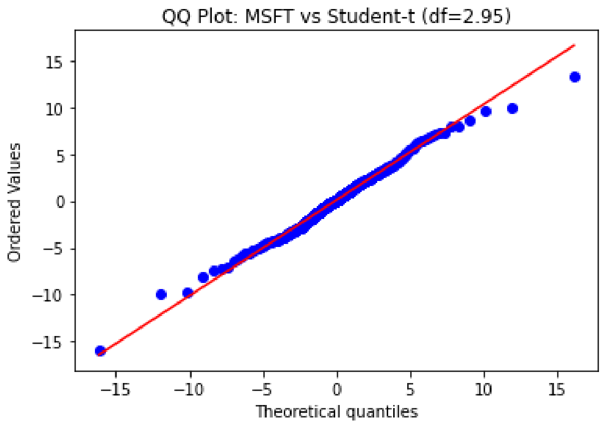
Figure 19: QQ plots (1–6) of assets vs Student-t



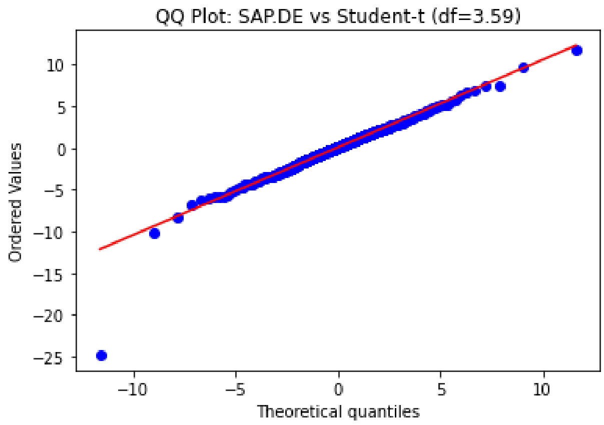
(a) AAPL



(b) ADS.DE



(c) MSFT



(d) SAP.DE

Figure 20: QQ plots (7–10) of assets vs Student-t