Statistical Methods for Green Stocks Analysis

Candidate Number: 39295, 34149, 34677, 45428

January 12, 2020

Abstract

Inspired by the Paris Agreement, we analysed some green stocks using statistical methods in this report. In PCA for the stock returns, the first principal component can be seen as the market factor and the second principal component could be the volatility against the market. In some green sectors such as solar energy, stock returns are highly correlated and extreme returns of different stocks usually coincide in addition to market crashes. The normality of the stock returns is always rejected, which also implies the non-normality of the loss distribution of portfolios. The return series are heavy-tailed and therefore the risk measures such as VaR and ES are typically underestimated using the normal distribution. The dependence structure between log returns can be modelled by different copulas and the best-fitting copulas in practice are Gauss, t and Frank copulas in most cases. The Marshall-Olkin Copula is also studied but it is not well-fitting in this scenario.

1 Introduction

The world is becoming green, from plastic recycling and sustainable power generation to organic groceries and temporarily closed fishing. The Paris agreement [1], which had been signed in Dec 2015, and an overall concern for climate change is having a significant impact in the different industries. More and more businesses were going "green" to obtain the many benefits from the agreement and the government. These benefits contain not only tax reductions and government subsidies but also the increasing demand for green stocks from various investors.

The target of this project is to analyze some green stocks together with the S&P500 index using different statistical methods. In Section 2, we make a brief introduction of the green stocks and their log returns. In Section 3, we study the empirical loss distribution and risk measures of different portfolios. In Section 4, we analyze the dependence structure of different stock pairs using various copulas and discuss the fittings. In Section 5, we apply the principal component analysis and also analyze the dependence structure between principal components and S&P500 index. In Section 6, we study the Marshall-Olkin copula and try to fit it to the data.

2 Data and Background

The basic information of the companies is shown in Table 1 and the log-returns for these 10 stocks and S&P500 are shown in Figure 1.

Table 1: Information of the selected companies

Ticker	Company Name	Industry	Market Cap[1]	Beta
GE	General Electric	Wind Power	65.84	1.01
ITT	ITT Industries	Water Purification	4.23	1.22
PCG	PG&E Corp	Hydro Operations	12.32	0.52
IDA	Idacorp	Hydro Operations	4.69	0.64
CVA	Covanta Holding Corp	Waste Reduction	1.76	0.80
UNFI	United Natural Foods	Organics	0.53	1.02
ORA	Ormat Technologies Inc.	Geothermal	2.65	1.04
SPWR	Sunpower Corp	Solar Energy	0.70	1.78
FSLR	First Solar	Solar Energy	4.45	1.56
TSLA	Tesla	Green Transportation	57.15	1.31

[1] Market capitalisation is calculated on 2018-12-31 and shown in billions.

All the stocks are correlated to some extent due to the factors that affect the whole market. For example on August 8, 2011, also known as Black Monday 2011, US stock markets crashed following the first time in history the United States sovereign debt was downgraded. The S&P500 index plunged 6.6% on that day and highly volatile after the crash. During the period between June 2015 and June 2016, investors sold shares globally due to the slowing growth in the GDP of China, which made a downturn in the market and S&P500 index. On February 5, 2018, the S&P500 index also went down 4.1% for market correlation.

Besides the market turbulence, extreme stock returns also coincide in sectors such as Solar Energy. For example on April 9, 2013, First Solar (FSLR), the world's largest thin-film solar manufacturer, soared 46% after forecasting yearly sales that would exceed estimates. The solar power plants maker, Sunpower Corp (SPWR) had a 17% rise on the same day. Additionally, extreme returns tend to cluster together, which is also known as the volatility clustering. For example, on November 14, 2018, PG&E (PCG) shares tumbled 21.8% and had lost nearly half of its value as California wildfire risks mounted. On the second day, PCG continued to fall 30.7% but recovered by 37.5% on the third day.

Based on the Paris agreement signed in December 2015, we expect upward movements of green stocks since government subsidies would directly result in cost reduction and profit improvement. However, it was reported that the positive effects brought by the financial benefits were not as significant as expected.

3 Loss and Risk Measures of the Portfolio

To study the distribution of total loss, three portfolios are constructed. The portfolio 1, 2 and 3 consist of the first 10, 5 and 3 stocks in the list respectively and the investments in each stock is $1000 during the

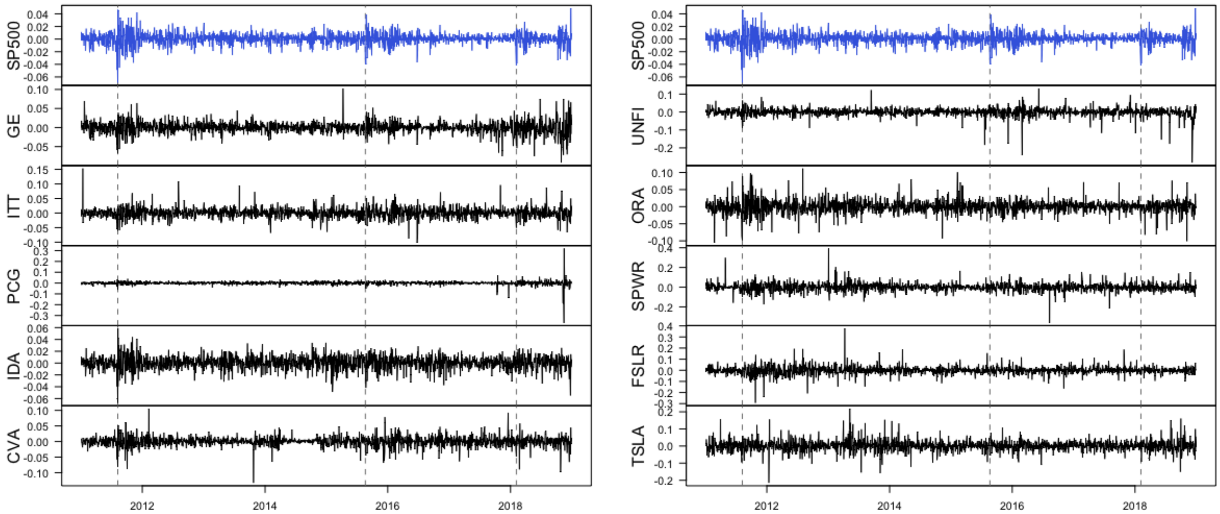

Figure 1: Log returns for 10 stocks with S&P500

entire period. Using the historical return series of 2012 sample points for each stock, the time series of the total loss of each portfolio can be computed for the period.

The Jarque-Bera test rejects the normality of the derived losses for all portfolios based on the small p-value and it is also justified by the QQ plots shown in 2. The Mardia test rejects the multivariate normality of the log returns of the stocks in all portfolios based on the small p-value. Therefore it can be concluded that the total losses of portfolios are not normally distributed as a result of the non-normality of log returns.

Table 2: Jarque-Bera test for losses of portfolios and Mardia test for log returns of portfolios

JB test	**pf 1**	**pf 2**	**pf 3**
p-value	0	0	0

Mardia test	**pf 1**	**pf 2**	**pf 3**
p-value(b,k)	(0,0)	(0,0)	(0,0)

The risk measures including value at risk (VaR) and expected shortfall (ES) are computed based on the empirical losses. Together with the empirical distribution of losses, the VaR and ES are plotted at level 95%, 97.5%, 99%, 99.5%, 99.9% and 99.99% from left to right, where the red dashed line and the black dashed line represents the risk measures with normal distribution and historical simulation (HS) respectively.

Comparing the black and red dashed lines in Figure 3, it is clear that for all portfolios, the normal distribution leads to smaller VaR and ES at all levels from 99% than the empirical distribution. The reason is that the normal distribution is not able to describe the heavy-tail of a probability distribution. Therefore the normal distribution would generally underestimate the risk in practice.

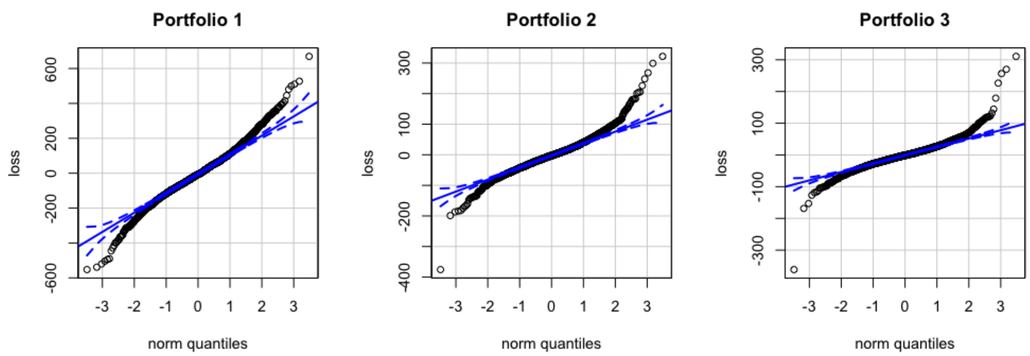

Figure 2: Normal QQ plots of losses for portfolios

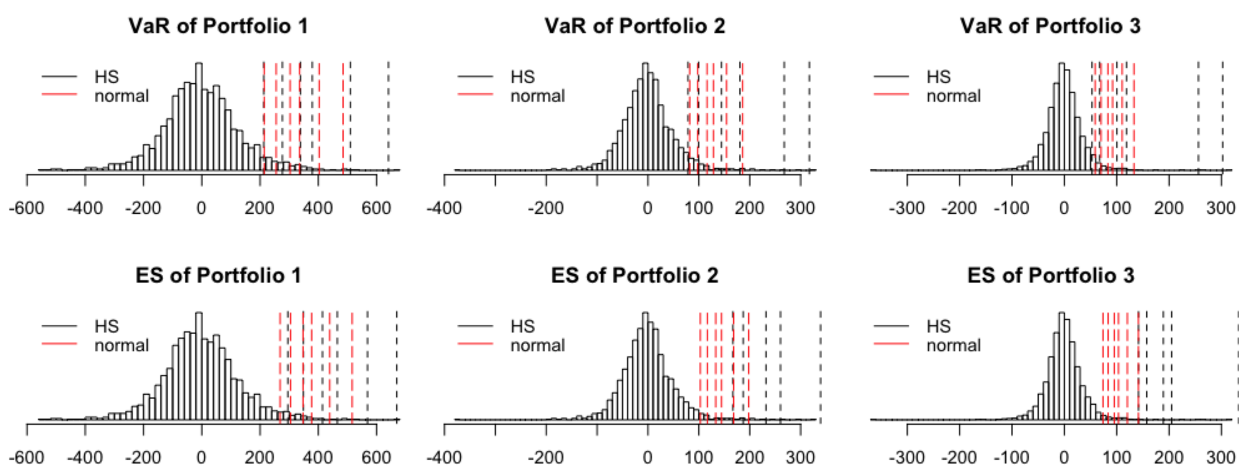

Figure 3: VaR(top) and ES(bottom) with empirical losses for portfolios

4 Dependence Structure

4.1 Methodology

In order to fit copulas to data, the function `fitCopula()` in package `copula` in R is used. The parameter `copula` defines the type of copula and the parameter `method` defines the estimation method. For fit diagnostics, the goodness-of-fit test and BIC are used via function `gofCopula()` and `BIC()`.

4.1.1 Copula

Five copulas are considered in this report including Gauss copula, Gumbel copula, t copula, Clayton copula and Frank copula. Gauss copula is a traditional copula with moderate tails for modelling dependence but may fail in some extreme events. Gumbel copula can be directly related to multivariate extensions of extreme value theory. Moreover, t copula can capture dependence in the tails while keeping the flexibility. Clayton copula is an asymmetric Archimedean copula which has a heavy density in the lower tail and an expanding cloud. Frank copula has limited concentrations and less probability concentrated in the tails than the Gumbel's. The copulas function are shown in Table 3.

Table 3: Summary of copulas used in this report

copula	**code**	**definition**
Gauss	`normalCopula()`	$C_P^{\text{Ga}}(u_1, u_2) = \Phi_P(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$
t	`tCopula()`	$C_{\nu,P}^{t}(u_1, u_2) = t_{\nu,P}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2))$
Clayton	`claytonCopula()`	$C_\theta^{\text{Cl}}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$
Gumbel	`gumbelCopula()`	$C_\theta^{\text{Gu}}(u_1, u_2) = \exp\left\{ -\left((-\log u_1)^\theta + (-\log u_2)^\theta\right)^{1/\theta}\right\}$
Frank	`frankCopula()`	$C_\theta^{\text{Fr}}(u_1, u_2) = -\frac{1}{\theta}\log\left(1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right)$

4.1.2 Estimation Method

In this section, pseudo maximum likelihood estimation and method of moments using rank correlation are used for each copula. Assume there exists samples $X_1, \ldots, X_n$. For pseudo MLE, one can firstly estimate the marginal distribution by empirical distribution $\hat{F}_i(x) = \frac{1}{n+1}\sum_{j=1}^n \mathbb{1}_{X_{ij}<x}$. Then the estimator $\hat{\theta}$ is obtained by maximising $\ln L = \sum_{j=1}^n \ln(c(\hat{F}_1(X_{1j}), \ldots, \hat{F}_d(X_{dj})|\theta))$. For method of moments, the inversion of Spearman's rho estimator is used for Gauss copula and the inversion of Kendall's tau estimator is used for t, Frank, Gumbel and Clayton copulas. These relations are shown in Table 4, where $D_1(.)$ is called the Debye function defined by $D_1(\alpha) = \frac{1}{\alpha}\int_0^\alpha \frac{t}{\exp(t)-1}dt$.

Table 4: Summary of relations used in the method of moments

	rho-Gauss	**tau-t**	**tau-Frank**	**tau-Clayton**	**tau-Gumbel**
equation	$\frac{6}{\pi}\sin^{-1}(\frac{\rho_{ij}}{2})$	$\frac{2}{\pi}\sin^{-1}(\rho_{ij})$	$\frac{\theta}{(\theta+2)}$	$1 - \frac{1}{\theta}$	$1 + \frac{4}{\theta}(D_1(\theta) - 1)$
code	`‘‘irho’’`	`‘‘itau.mpl’’`	`‘‘itau’’`	`‘‘itau’’`	`‘‘itau’’`

4.1.3 Copula Selection Criterion

In order to select the best copula, Bayesian information criterion (BIC) is introduced to compare copula models. BIC is defined as follows,

$$\text{BIC} = \log(n)k - \log(L) \tag{1}$$

where k is the number of parameters in the model, n is the sample size and L is the maximised log-likelihood. As suggested by Equation (1), the BIC introduces the number of parameters k as a penalty term based on the sample size. The copula with the lowest BIC would be the best-fitting copula as it makes a balance between the least parameters and the highest log-likelihood.

Besides, the goodness-of-fit test [2] is used to compare the non-parametric copula with a parametric family of copulas. The null hypothesis $H_0 : C \in \{C_\theta\}$ is that the copula C comes from a parametric family of copulas whose members are defined by a parameter θ. The Cramer-von Mises statistic for the goodness-of-fit test is defined as,

$$S_n = n \int_{[0,1]^d} \left\{ \hat{C}_n(\mathbf{u}) - C_{\theta_n}(\mathbf{u}) \right\}^2 d\hat{C}(\mathbf{u}) = \sum_{i=1}^{n} \left\{ \hat{C}_n(\hat{\mathbf{u}}) - C_{\theta_n}(\hat{\mathbf{u}}) \right\}^2$$

where the empirical copula of n independent observations $\hat{C}_n$ is the estimator of C and θ_n the estimator of θ. The parametric bootstrap procedure [3] is suggested and an approximate p-value is given by

$$\frac{1}{N} \sum_{k=1}^{N} \mathbb{1}(S_n^{(k)} \geq S_n)$$

where N is the number of bootstrap and $S_n^{(k)}$ the Cramer-von Mises statistic in the k-th bootstrap.

In this report, a significance level of 95% is used so that the null hypothesis is rejected for p-value ≤ 0.05. Note that for the t copula, the degrees of freedom are not considered as a parameter to be estimated, which implies the limitation that the test for t copula may not be reliable.

4.2 Sample Rank Correlation

As shown in Table 5, the sample Spearman's rho and Kendall's tau for GE and ITT imply a moderately positive correlation between these two companies. The correlation of SPWR and FSLR is stronger than the previous pair. This can be due to the fact that SPWR and FSLR are both the rapidly developing sector of solar energy. In contrast, ITT is a water purification company but GE's specialism is wind power. A weak correlation is shown between FSLR and TSLA. It is partly because TSLA is a company that sells electric cars across the world but FSLR is a company that sells solar panel with photovoltaic cells on its surface. Comparing to the other 3 pairs, one can conclude that PCG and SPWR are nearly independent based on the small rank correlations.

Table 5: Sample Spearman's rho and Kendall's tau for different pairs

	GE-ITT	**SPWR-FSLR**	**FSLR-TSLA**	**PCG-SPWR**
rho	0.471	0.666	0.286	0.0999
tau	0.333	0.488	0.195	0.0665

4.3 Copula Fitting

To analyse the dependence structure of pairs GE-ITT, SPWR-FSLR, FSLR-TSLA and PCG-SPWR, different types of copulas are implemented including Gauss, t, Frank, Gumbel and Clayton copulas. The pseudo maximum likelihood estimation and method-of-moments are used when fitting copulas to data. The goodness-of-fit test is applied to each type of copula and the copulas that reject the null hypothesis with a p-value less than 0.05 is not considered as well-fitting copulas. The best-fitting copula is selected based on the lowest BIC among the rest of copulas.

The summary of results are shown in Table 6 and 7. Note that the Clayton copula is not applicable for FSLR-TSLA and PCG-SPWR due to the limitation that the loglikelihood goes to negative infinity when the parameter θ is large.

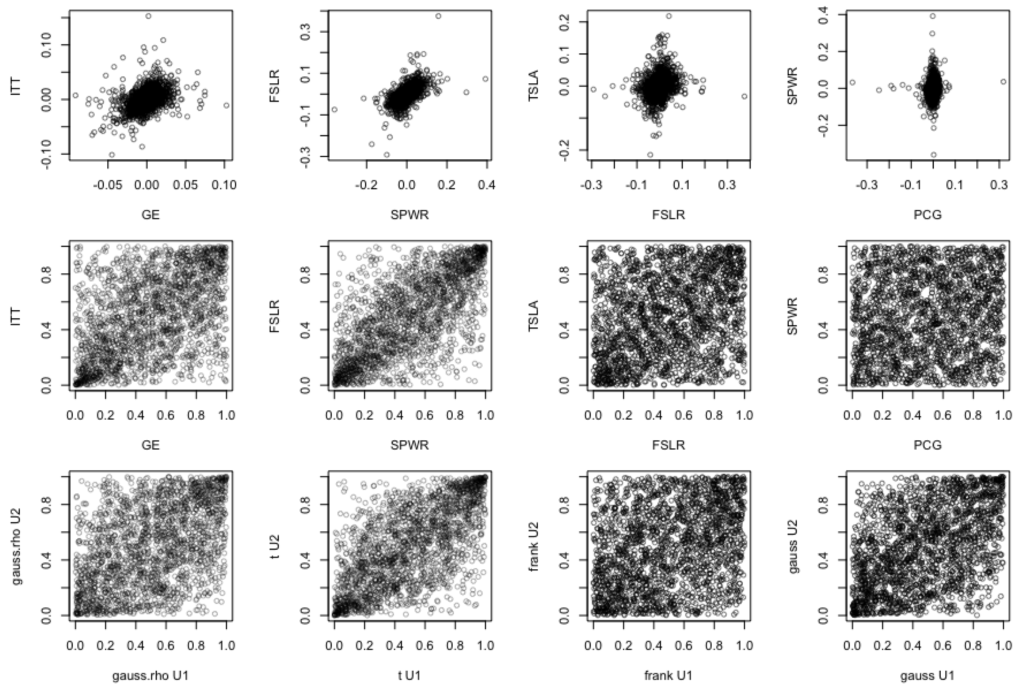

Figure 4: Scatter plots of log returns(top), pseudo observations(middle) and simulations(bottom) for GE-ITT, SPWR-FSLR, FSLR-TSLA and PCG-SPWR

By combining the p-value obtained from the goodness-of-fit tests and the BIC values, it can be shown that the Gaussian copula estimated using the sample Spearman's rho (gauss.rho) is acceptable for the pair of GE and ITT, which is determined by the fact that p-value > 0.05. It is recorded that the Spearman's rho of gauss.rho is the same as the sample one (0.471) and Kendall's tau of gauss.rho is almost the same as the sample one (0.325). The lower and upper tail dependence of the gauss.rho are both consistent with the first bottom plot.

For the pair of SPWR and FSLR, gauss.rho, t copula estimated by pseudo MLE (t) and t copula estimated by inversion of sample tau (t.tau) are viable in the goodness-of-fit test. Among these 3 remaining copulas, the t copula is the most suitable for SPWR and FSLR because this copula has the lowest BIC values, in comparison to other two copulas. Also, the Spearman's rho and Kendall's tau of t copula are very similar to the sample ones, which are 0.670 and 0.488 respectively. It is clear that the lower and upper tail dependence of the second bottom plot are greater than 0, which coincides with the fact that t copula has non-zero lower and upper tail dependence (0.281).

Table 6: Summary of fits for pairs GE-ITT and SPWR-FSLR

GE-ITT	**gof p-value**	**BIC**	**loglik**	ρ_S	ρ_τ	λ_L	λ_U
gauss	0.00249	-488	248	0.452	0.311	0	0
gauss.rho	0.769	-486	247	0.471	0.325	0	0
t	0.00746	-574	295	0.478	0.33	0.206	0.206
t.tau	0.00249	-574	295	0.481	0.333	0.207	0.207
frank	0.00249	-517	263	0.491	0.339	0	0
frank.tau	0.00249	-517	262	0.483	0.333	0	0
gumbel	0.00249	-446	227	0.427	0.296	0	0.372
gumbel.tau	0.00249	-437	222	0.474	0.333	0	0.412
clayton	0.00249	-501	254	0.409	0.281	0.412	0
clayton.tau	0.00249	-475	241	0.477	0.333	0.499	0
SPWR-FSLR	**gof p-value**	**BIC**	**loglik**	ρ_S	ρ_τ	λ_L	λ_U
gauss	0.00249	-1177	593	0.651	0.466	0	0
gauss.rho	0.903	-1175	591	0.666	0.479	0	0
t	0.0871	-1235	625	0.666	0.479	0.281	0.281
t.tau	0.619	-1234	625	0.677	0.488	0.286	0.286
frank	0.00249	-1195	601	0.686	0.492	0	0
frank.tau	0.00249	-1195	601	0.681	0.488	0	0
gumbel	0.00249	-1158	583	0.63	0.455	0	0.541
gumbel.tau	0.00249	-1146	577	0.67	0.488	0	0.574
clayton	0.00249	-893	450	0.525	0.37	0.554	0
clayton.tau	0.00249	-689	348	0.67	0.488	0.696	0

Similarly, by inspection of the corresponding p-value and BIC values of FSLR and TSLA, frank copula is the most suitable copula for FSLR and TSLA. The Spearman's rho (0.280) and Kendall' tau (0.185) are very close to the sample ones. This indicates a weak correlation between FSLR and TSLA. The lower and upper tail dependence of the frank copula are both 0, which is consistent with the third bottom plot.

But it turns out to be a different case to PCG and SPWR for which the best model is gauss. Its corresponding Spearman's rho and Kendall's tau are 0.103 and 0.0686 respectively. This tells us that PCG and SPWR are nearly independent of each other. The lower and upper tail dependence of gauss is both 0, which is the same scenario as our first pair of GE and ITT.

The scatter plots of data simulated by these estimated best-fitting copulas are shown in Figure 4 together with the log returns and pseudo observations. For PCG and SPWR, the plots imply that the estimated copula may overestimate the dependence level between the stocks.

5 Principal Component Analysis

Principal Component Analysis (PCA) is a dimensionality reduction method that selects a set of uncorrelated linear combinations called principal components which account for the most variability of the original data. In this case, PCA is applied to the log returns of 10 green stocks and new indexes are constructed on the first and second principal components. Furthermore, the dependence structures of these indexes and S&P500 are modelled by different types of copulas.

Table 7: Summary of fits for pairs FSLR-TSLA and PCG-SPWR

FSLR-TSLA	**gof p-value**	**BIC**	**loglik**	$\boldsymbol{\rho}_S$	$\boldsymbol{\rho}_\tau$	$\boldsymbol{\lambda}_L$	$\boldsymbol{\lambda}_U$
gauss	0.0423	-153	80.3	0.267	0.18	0	0
gauss.rho	0.938	-152	79.8	0.286	0.193	0	0
t	0.112	-152	83.7	0.275	0.185	0.00612	0.00612
t.tau	0.415	-152	83.5	0.289	0.195	0.00739	0.00739
frank	0.0622	-168	87.7	0.293	0.198	0	0
frank.tau	0.296	-168	87.7	0.289	0.195	0	0
gumbel	0.00249	-122	64.8	0.24	0.161	0	0.212
gumbel.tau	0.00249	-116	61.7	0.288	0.195	0	0.253
clayton	NA	NA	NA	NA	NA	NA	NA
clayton.tau	0.00249	-115	61.2	0.288	0.195	0.239	0
PCG-SPWR	**gof p-value**	**BIC**	**loglik**	$\boldsymbol{\rho}_S$	$\boldsymbol{\rho}_\tau$	$\boldsymbol{\lambda}_L$	$\boldsymbol{\lambda}_U$
gauss	0.759	-15.5	11.5	0.103	0.0686	0	0
gauss.rho	0.993	-15.5	11.5	0.0999	0.0667	0	0
t	0.799	-7.65	11.4	0.103	0.0686	1.31e-14	1.31e-14
t.tau	0.759	-7.86	11.5	0.0996	0.0665	0	0
frank	0.629	-12.5	10	0.0996	0.0665	0	0
frank.tau	0.863	-12.5	10	0.0996	0.0665	0	0
gumbel	0.0373	-8.42	8.01	0.0716	0.0493	0	0.0671
gumbel.tau	0.102	-6.81	7.21	0.0974	0.0665	0	0.0901
clayton	NA	NA	NA	NA	NA	NA	NA
clayton.tau	0.555	-12.1	9.87	0.0995	0.0665	0.00771	0

5.1 Principal Components

The first and second principal components explain 33.0% and 12.8% of the total variance respectively. They explain almost half of the total variance.

In the direction of the first principal component (Dim1), all the stocks have the same signs and all the coefficients are positive. This implies that the first component can be seen as the systematic factor or the market factor. It is composed of macroeconomic factors such as fiscal and monetary policy as well as international regulation, which can make an impact on all the stocks in the market. This is the reason why all the stocks have similar contributions to the first component as shown in Figure 5.

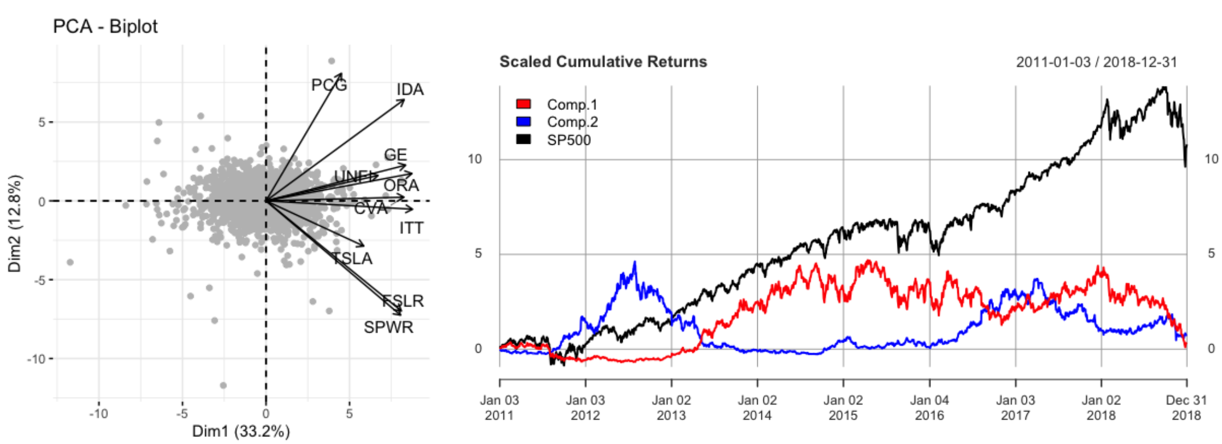

Figure 5: Biplot of PCA(left) and scaled cumulative returns of Comp.1, Comp.2 and S&P500(right)

In the direction of the second principal component (Dim2), stocks have both positive and negative signs, where PCG and IDA have the greatest positive coefficients and FSLR and SPWR have the greatest negative coefficients. The main information of them is shown in Table 8. While PCG and IDA have low beta and standard deviation of log returns, FSLR and SPWR have higher volatility. Besides, both PCG and IDA are from hydro operation industry and both FSLR and SPWR are from the Solar Energy industry. It is notable that PCG and IDA are local electricity monopolies whereas the SPWR and FSLR are competing with solar panel makers globally. In general, the monopoly companies have lower volatility in relation to the market based on the long-term customer relationship and customer loyalty. Therefore the second component is supposed to indicate the inverse volatility in relation to the market or the monopoly level.

Table 8: Features of PCG, IDA, FSLR and SPWR

Group	**Ticker**	**Coefficient**	**Beta**	**Stdev**	**Industry**
A	PCG	0.536	0.52	0.0185	Hydro Operations
	IDA	0.426	0.64	0.0110	Hydro Operations
B	FSLR	-0.463	1.56	0.0351	Solar Energy
	SPWR	-0.481	1.78	0.0401	Solar Energy

where “Stdev” represents the standard deviation.

The scaled cumulative returns of the first and second principal components are plotted in Figure 5 together with S&P500. It is clear that the first component and S&P500 have a similar trend during the whole period while the first and second components move in the opposite direction. This also verifies that the first component is the market factor.

There are some limitations of PCA. For example, PCA relies on linear assumptions and is not able to find non-linear relationships in the data. Furthermore, PCA may lose some of the information due to orthogonal assumptions.

5.2 Dependence Structure

As shown in the scatter plots of log returns and the pseudo observations in Figure 6, it is clear that the first principal component and S&P500 are positively correlated while the second principal component and S&P500 are weakly negatively correlated. Note that the parameter of Gumbel copula has restriction $\theta \geq 1$ and the boundary case is always optimal for negatively correlated data. Therefore in order to apply different types of copula including Gumbel copula, the second principal component is rotated 90 degrees.

For Comp.1 and S&P500, the goodness-of-fit test shows that all the copulas are not well-fitting except for the Gauss copula estimated by the sample Spearman’s rho (Gauss.rho). Hence it is considered as the best-fitting copula even though the t copulas have lower BIC and higher loglikelihood. The scatter plot of simulations of the estimated copula is shown in the top right of Figure 6. Since it is estimated by the inversion of sample rho, it has the same Spearman’s rho as the sample (0.739) and also similar Kendall’s tau (0.544), which shows a strong positive rank correlation between Comp.1 and S&P500. However, the Gauss copula has zero lower and upper tail dependence by definition, which may not comply with the scatter plot

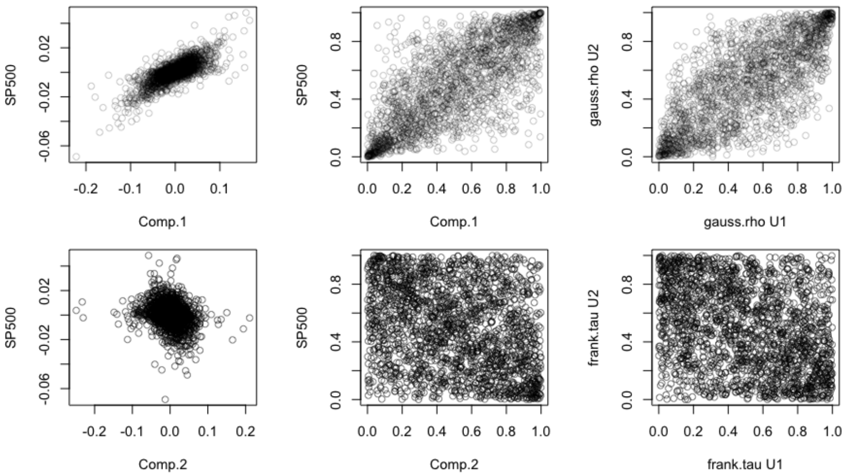

Figure 6: Scatter plots of log-returns(left), pseudo observations(middle) and simulations(right) for Comp.1-SP500(top) and Comp.2-SP500(bottom)

of pseudo observations in Figure 6, where a clear lower tail dependence can be seen in the bottom left corner.

Table 9: Summary of fits for Comp.1 and S&P500 as well as the rotated Comp.2 and S&P500

Comp.1-SP500	**gof p-value**	**BIC**	**loglik**	$\boldsymbol{\rho}_S$	$\boldsymbol{\rho}_\tau$	$\boldsymbol{\lambda}_L$	$\boldsymbol{\lambda}_U$
gauss	0.00249	-1744	876	0.748	0.553	0	0
gauss.rho	0.813	-1742	875	0.739	0.544	0	0
t	0.00746	-1840	928	0.751	0.555	0.426	0.426
t.tau	0.00249	-1840	928	0.75	0.555	0.426	0.426
frank	0.00249	-1599	803	0.755	0.555	0	0
frank.tau	0.00249	-1599	803	0.755	0.555	0	0
gumbel	0.00249	-1620	814	0.712	0.527	0	0.612
gumbel.tau	0.00249	-1609	808	0.742	0.555	0	0.638
clayton	0.00249	-1620	814	0.668	0.487	0.694	0
clayton.tau	0.00249	-1531	769	0.741	0.555	0.757	0
Comp.2.rot-SP500	**gof p-value**	**BIC**	**loglik**	$\boldsymbol{\rho}_S$	$\boldsymbol{\rho}_\tau$	$\boldsymbol{\lambda}_L$	$\boldsymbol{\lambda}_U$
gauss	0.00746	-168	88	0.279	0.188	0	0
gauss.rho	0.888	-168	87.8	0.29	0.196	0	0
t	0.0373	-165	90	0.283	0.191	0.0019	0.0019
t.tau	0.0473	-165	89.9	0.292	0.197	0.0021	0.0021
frank	0.0373	-172	89.7	0.296	0.199	0	0
frank.tau	0.137	-172	89.7	0.292	0.197	0	0
gumbel	0.00249	-121	64.5	0.24	0.162	0	0.212
gumbel.tau	0.00249	-115	61.1	0.291	0.197	0	0.255
clayton	NA	NA	NA	NA	NA	NA	NA
clayton.tau	0.00249	-149	78.3	0.291	0.197	0.244	0

For the rotated Comp.2 and S&P500, the Frank copula estimated by the sample Kendall's tau (Frank.tau) is considered as the best fitting copula based on the lowest BIC and the p-value of the goodness-of-fit test. The scatter plot of simulations of the estimated copula is shown in the bottom right of Figure 6. It has the same Kendall's tau as the sample (0.197) and also similar Spearman's rho (0.292), which shows a weak positive rank correlation between the rotated Comp.2 and S&P500 and hence a weak negative rank correlation between Comp.2 and S&P500. Moreover, Frank copula has zero lower and upper tail dependence, which is consistent with the scatter plot as shown in Figure 6.

6 Marshall-Olkin Copula

In this section, we wish to describe the theoretical aspects of Marshall-Olkin Copula and present the fitting procedure of this copula to the available data.

In this project, we restrict ourselves to only discussing bivariate Marshall-Olkin copula. For further details about n-dimensional case and Marshall-Olkin distributions, please refer to Marshall and Olkin [4], and Muliere and Scarsini [5].

6.1 Bivariate Marshall-Olkin Copula

Let us consider a two-component system where the components are subject to shocks, which are fatal to one or both components. Consider X and Y as the lifetimes of the two components. Moreover, assume that the shocks follow three independent Poisson processes with parameters $\lambda_1, \lambda_2, \lambda_3 \geq 0$. where the index indicates whether the shocks influence only component 1, only component 2 or both. Then the times Z_1, Z_2 and Z_3 of occurrence of these shocks are independent exponential random variables with parameters λ_1, λ_2 and λ_3, respectively. Therefore,

$$\bar{F}(x,y) = \mathbb{P}(X > x, Y > y) = \mathbb{P}(Z_1 > x)\mathbb{P}(Z_2 > y)\mathbb{P}(Z_3 > \max(x,y)) \tag{2}$$

As X and Y have univariate (marginal) survival functions $\bar{F}_1(x) = \exp\big(-(\lambda_1+\lambda_3)x\big)$ and $\bar{F}_2 = \exp\big(-(\lambda_2+\lambda_3)y\big)$, and $\max(x,y) = x + y - \min(x,y)$, then

$$\begin{aligned}\bar{F}(x,y) &= \exp\big(-(\lambda_1+\lambda_3)x - (\lambda_2+\lambda_3)y + \lambda_3 \min(x,y)\big)\\ &= \bar{F}_1\bar{F}_2 \min\big(\exp(\lambda_3 x), \exp(\lambda_3 y)\big)\end{aligned} \tag{3}$$

Consider $\alpha = \lambda_3/(\lambda_1+\lambda_3)$ and $\beta = \lambda_3/(\lambda_2+\lambda_3)$, hence $\bar{F}_1(x)^{-\alpha} = \exp(\lambda_3 x)$ and $\bar{F}_2(x)^{-\beta} = \exp(\lambda_3 y)$ and the survival copula of $(X_1, X_2)^T$ is as follows

$$\hat{C}(u,v) = uv\min(u^{-\alpha}, v^{-\beta}) = \min(u^{1-\alpha}v, uv^{1-\beta}) \tag{4}$$

The survival copula $\hat{C}$ for the Marshall-Olkin bivariate family distribution leads to a copula with two parameters such that

$$C(u,v) = \min(u^{1-\alpha}v, uv^{1-\beta}) = \begin{cases} u^{1-\alpha}v, & \text{if } u^\alpha \geq v^\beta \\ uv^{1-\beta}, & \text{if } u^\alpha \leq v^\beta \end{cases}, \quad 0 \leq \alpha, \beta \leq 1 \tag{5}$$

This is particularly known as the *Marshall-Olkin family*. The density function c generated by its copula C is given by

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v} = \begin{cases} (1-\alpha)u^{-\alpha}, & \text{if } u^\alpha > v^\beta \\ (1-\beta)v^{-\beta}, & \text{if } u^\alpha < v^\beta \end{cases} \tag{6}$$

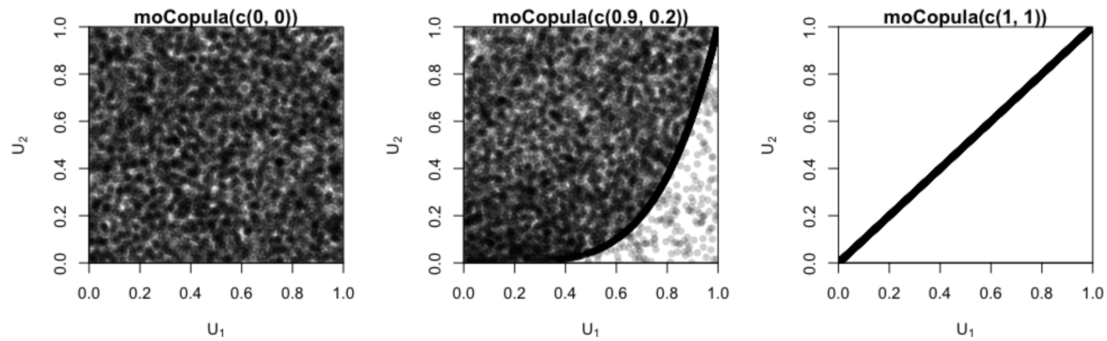

Figure 7: Scatter plots of Marshall-Olkin copula where
$\alpha = \beta = 0$(left), $\alpha = 0.9, \beta = 0.2$ (middle), $\alpha = \beta = 1$ (right)

As shown in Figure 7 for different pairs of λ_1 and λ_2, the mass of singular component is concentrated on the curve $u^\alpha = v^\beta$ in $[0, 1]^2$. For the boundary cases, as $\alpha, \beta \to 0$, $C(u, v) = uv$ becomes the independent copula. As $\alpha, \beta \to 1$, $C(u, v) = \min(u, v)$ becomes the comonotone copula. Then we can derive the Kendall's tau and Spearman's rho as follows.

$$\rho_S = 12 \int_0^1 \int_0^1 C(u,v) du dv - 3 = 12 \int_0^1 \Big(\int_0^{u^{\alpha/\beta}} u^{1-\alpha} v dv + \int_{u^{\alpha/\beta}}^1 u v^{1-\beta} dv \Big) du - 3 = \frac{3\alpha\beta}{2\alpha + 2\beta - \alpha\beta}$$

$$\rho_\tau = 4 \int_0^1 \int_0^1 C(u,v) dC(u,v) - 1 = 4 \Big(\frac{1}{2} - \int_0^1 \int_0^1 \frac{\partial}{\partial u} C(u,v) \frac{\partial}{\partial v} C(u,v) du dv \Big) - 1 = \frac{\alpha\beta}{\alpha + \beta - \alpha\beta}$$

We can derive the coefficient for lower tail and upper tail dependence as follows.

$$\lambda_L = \lim_{u \to 0} \frac{C(u,v)}{u} = \lim_{u \to 0} \frac{\min(u^{2-\alpha}, u^{2-\beta})}{u} = \lim_{u \to 0} (2 - \min(\alpha, \beta)) u^{1 - \min(\alpha,\beta)} = 0$$

$$\lambda_U = \lim_{u \to 1} \frac{\hat{C}(1-u, 1-u)}{1-u} = \lim_{u \to 1} \frac{1 - 2u + C(u,u)}{1-u} = 2 - \lim_{u \to 1} (2 - \min(\alpha, \beta)) u^{1 - \min(\alpha,\beta)} = \min(\alpha, \beta)$$

6.2 Copula Fitting

In order to estimate the parameters α and β in Marshall-Olkin copulas, the pseudo MLE is used based on the copula density $c(u, v)$ derived in Equation (6).

Algorithm to fit Marshall-Olkin copula, The following algorithm estimate the parameters α and β of Marshall-Olkin copula $C(u, v)$:

Step 1. Estimate the marginal distribution by empirical distribution $\hat{F}_X, \hat{F}_Y$.

Step 2. Under constraints $0 \le \alpha, \beta \le 1$, estimate ML estimators $\hat{\alpha}$ and $\hat{\beta}$ by maximising

$$\ln L(\alpha, \beta) = \sum_{i=1}^n \ln(c(\hat{F}_X(x_i), \hat{F}_Y(y_i) | \alpha, \beta))$$

By definition, Marshall-Olkin copulas model positive random variables X and Y. Since the log returns are not strictly positive, one may use the exponential of the log returns. As a monotonically increasing function, the exponential transformation would not change the empirical marginal distribution $\hat{F}_X, \hat{F}_Y$. Therefore as shown in Figure 8, fitting copulas for log returns X is equivalent to fitting copulas of the exponential for log returns $\exp(X)$.

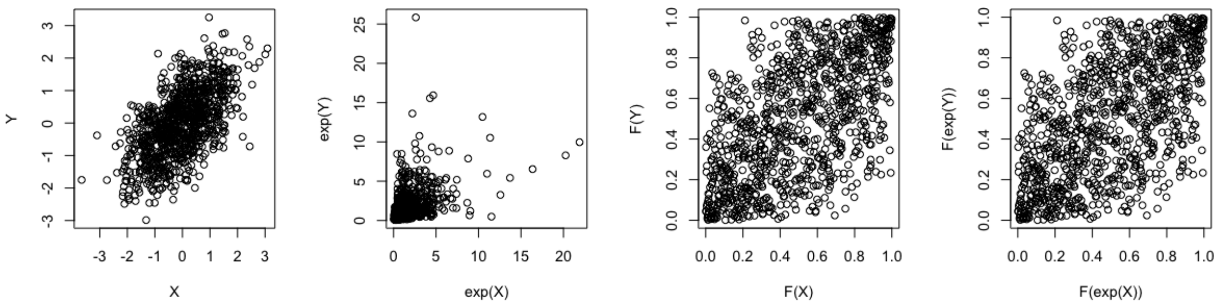

Figure 8: Exponential transformation of multivariate normal simulations

The results of Marshall-Olkin copula estimation for different pairs mentioned in previous sections are shown in Table 10. It is notable that all the pairs optimise beyond the constraints $0 \leq \alpha, \beta \leq 1$, which may imply that the Marshall-Olkin copula is not a well-fitting copula for these data.

Table 10: Summary of fits

	GE-ITT	**SPWR-FSLR**	**FSLR-TSLA**	**PCG-SPWR**	**Comp.1-SP500**	**Comp.2-SP500**
α	0.02772	0.04704	0.01853	-0.00197	0.05634	0.00723
β	-0.00198	-0.00188	-0.00188	0.01714	-0.00191	-0.00191

7 Conclusion

Inspired by the Paris Agreement, we analysed some green stocks using statistical methods in this report. In PCA for the stock returns, the first principal component can be seen as the market factor and the second principal component could be the volatility against the market. In some green sectors such as solar energy, stock returns are highly correlated and extreme returns usually coincide in addition to market crashes. The normality of the stock returns is always rejected, which also implies the non-normality of the loss distribution of portfolios. The return series are heavy-tailed and therefore the risk measures such as VaR and ES are typically underestimated using the normal distribution. The dependence structure between log returns can be modelled by different copulas and the best-fitting copulas in practice are Gauss, t and Frank copulas in most cases. The Marshall-Olkin Copula is also studied but it is not well-fitting in this scenario.

There are some limitations to the methods in this report. One limitation of PCA is the linear and orthogonal assumptions, which make it unable to find non-linear relationship. To move beyond linearity, one can consider an extension of PCA called kernel PCA using techniques of kernel methods. For copulas, we only studied bivariate cases in this report where multivariate copulas are also applicable. We also found that Clayton and Gumbel's copulas are not well-fitting in most cases. Instead, one can use the mixes between Archimedian copulas such as Clayton-Gumbel copula and Joe-Clayton copula, which allows asymmetries in left and right tails by introducing more parameters.

References

[1] Paris Agreement. United Nations. *United Nations Treaty Collect*, pages 1–27, 2015.

[2] C. Genest, W. Huang, and J. Dufour. A regularized goodness-of-fit test for copulas. *Journal de la Société française de statistique*, 154:64–77, 2013.

[3] C. Genest and B. Rémillard. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. In *Annales de l'IHP Probabilités et statistiques*, volume 44, pages 1096–1127, 2008.

[4] A. W. Marshall and I. Olkin. A multivariate exponential distribution. *Journal of the American Statistical Association*, 62:30–44, 1967.

[5] P. Muliere and M. Scarsini. Characterization of a Marshall-Olkin type class of distributions. *Annals of the Institute of Statistical Mathematics*, 39:429–441, 1987.

Code

```
###' Code Contribution ##################################################################
###' 39295: d, e, f
###' 34677: c, e, f
###' 34149: a, b, c
###' 45428: a, b, c

### Package ##############################################################################
library(xts) # xts
library(car) # qqPlot
library(QRM) # ESnorm
library(copula)
library(factoextra) # plot for pca
library(tseries) # jarque.bera.test


### (a) Data #############################################################################
setwd("")
tmp = read.csv("429stocks.csv", stringsAsFactors=FALSE)
tmp$TICKER[tmp$TICKER=="SPWRA"] = "SPWR" # same company
# obtain tickers and dates
tickers = unique(tmp$TICKER)
dates = as.Date(as.character( unique(tmp$date) ), format="%Y%m%d")
# returns for stocks
RET = c()
for (t in tickers){
  RET = cbind(RET, tmp$RET[tmp$TICKER==t])
}
colnames(RET) = tickers
RET = xts(RET, order.by=dates)
# obtain marketcap ( price * outstanding shares )
tmp = tmp[tmp$date=="20181231",]
matrix(tmp$PRC*tmp$SHROUT, nrow=1, dimnames=list("marketcap", tickers))/1000000

# SP500 return
tmp = read.csv("429sp.csv", stringsAsFactors=FALSE)
SP500 = xts(tmp$vwretd, order.by=dates)
# obtain beta ( cov(r_i, r_market) / var(r_market) )
apply(RET,2,cov,y=SP500)/var(as.vector(SP500))

# log returns # RET_{t}:=S_{t+1}/S_{t}-1, X_{t}=log(S_{t+1}/S_{t})=log(RET_{t}+1)
RET = log(RET+1)
SP500 = log(SP500+1)



### (b) Log Returns ######################################################################
par(mfrow=c(1,1))
ylab = c("SP500", tickers[1:5], "SP500", tickers[6:10])
plot.zoo(cbind(SP500,RET[,1:5], SP500, RET[,6:10]), screens=1:12, col=c("royalblue",rep(1,5))
         , las=1, main="", xlab="", ylab=ylab, cex.main=1
         , panel=function(x,y,...){abline(v=as.Date(c("2011-08-08","2015-08-21","2018-02-05")), col="salmon",lty=2);lines(x, y,...)})



### (c) Portfolio ########################################################################

#' @title function for computing loss, normal VaR and ES, HS VaR and ES of a portfolio
#' @param RET vector or matrix, returns
#' @param weights vector, weights of each asset
#' @param value numeric, total portfolio value
#' @param alpha vector, significance level of VaR and ES
#' @param linear logical, indicator of whether the RET is linear-return or log-return
#' @return a list of loss, normal VaR and ES, HS VaR and ES
summary.portfolio = function(RET, pf.weights, pf.value, alpha, linear=FALSE){
  # loss
  loss = loss.portfolio(RET, pf.weights, pf.value, linear=linear)
  # normal VaR and ES
  mu.hat = colMeans(RET)
  sigma.hat = var(RET)
  meanloss = -sum(pf.weights*mu.hat) * pf.value
  varloss = pf.value^2 * as.numeric(t(pf.weights) %*% sigma.hat %*% pf.weights)
  VaR.normal = meanloss + sqrt(varloss) * qnorm(alpha)
  ES.normal = meanloss + sqrt(varloss) * dnorm(qnorm(alpha))/(1-alpha)
  # HS VaR and ES
  VaR.hs = quantile(loss,alpha)
```

```
  ES.hs = apply(as.array(VaR.hs), 1, function(VaR){ mean(loss[loss > VaR]) } )
  list(loss=loss, VaR.normal=VaR.normal, ES.normal=ES.normal, VaR.hs=VaR.hs, ES.hs=ES.hs)
}
#' function for computing total loss of a portfolio
loss.portfolio = function(RET, weights, value, linear=FALSE){
  if (linear) { RET = RET } else { RET = (exp(RET)-1) }
  Profit = RET * value * weights
  -rowSums(Profit)
}
#' function for plotting the histogram of loss with VaR or ES
plot.portfolio = function(portfolio, type=c("VaR","ES"),...){
  hist(portfolio$loss, nclass=50, prob=TRUE, xlab="Loss", ...)
  measure.hs = switch(type, VaR=portfolio$VaR.hs, ES=portfolio$ES.hs)
  measure.normal = switch(type, VaR=portfolio$VaR.normal, ES=portfolio$ES.normal)
  abline(v=measure.hs, col=1, lty=2)
  abline(v=measure.normal, col=2, lty=5)
  legend("topleft", legend=c("HS","normal"), col=1:2, lty=1, bty="n")
}

# compute loss, normal VaR and ES, HS VaR and ES
pf1 = summary.portfolio(RET, pf.weights=rep(0.1,10), pf.value=10000, alpha=alpha)
pf2 = summary.portfolio(RET[,1:5], pf.weights=rep(0.2,5), pf.value=5000, alpha=alpha)
pf3 = summary.portfolio(RET[,1:3], pf.weights=rep(1/3,3), pf.value=3000, alpha=alpha)
# check normality
par(mfrow=c(1,3), cex=.9)
qqPlot(pf1$loss, id=F, ylab="loss", main="Portfolio 1")
qqPlot(pf2$loss, id=F, ylab="loss", main="Portfolio 2")
qqPlot(pf3$loss, id=F, ylab="loss", main="Portfolio 3")
jarque.bera.test(pf1$loss)$p.value
jarque.bera.test(pf2$loss)$p.value
jarque.bera.test(pf3$loss)$p.value
MardiaTest(as.matrix(RET))
MardiaTest(as.matrix(RET[,1:5]))
MardiaTest(as.matrix(RET[,1:3]))
# histogram and comparing
par(mfrow=c(2,3), cex=.9)
par(mar = c(3.1, 1.1, 3.1, 0.1))
plot.portfolio(pf1, type="VaR", main="VaR of Portfolio 1", yaxt="n")
plot.portfolio(pf2, type="VaR", main="VaR of Portfolio 2", yaxt="n")
plot.portfolio(pf3, type="VaR", main="VaR of Portfolio 3", yaxt="n")
plot.portfolio(pf1, type="ES", main="ES of Portfolio 1", yaxt="n")
plot.portfolio(pf2, type="ES", main="ES of Portfolio 2", yaxt="n")
plot.portfolio(pf3, type="ES", main="ES of Portfolio 3", yaxt="n")



### (d) Copulas ##############################################################################

## (d.1) Pseudo-observations ------------------------------------
Uret = pobs(RET, ties.method = "max") # apply(RET, 2, edf, adjust=1)
# scatterplot
tickers = c("GE","ITT","PCG","SPWR","FSLR","TSLA")
pairs2(RET[,tickers], cex=0.3, col=adjustcolor("black",alpha.f=0.3))
pairs2(Uret[,tickers], cex=0.3, col=adjustcolor("black",alpha.f=0.3))


## (d.2) Sample Correlations ------------------------------------
cor(Uret[,tickers], method="spearman")
cor(Uret[,tickers], method="kendall")


## (d.3) Fit Copula --------------------------------------------

#' @title fit bivarite copulas including gauss,t,gumbel,clayton and frank copulas
#' @param U matrix, bivarite pseudo observations
#' @return list, consists of different fitCopula class objects
fit.copulas = function(U){
  if (ncol(U)!=2) stop("ncol(U) is not 2")
  # gauss copula
  fit.gauss = tryFitCopula(normalCopula(dim=2,dispstr="un"), data=U, method="mpl")
  # gauss copula - method of moments: P^ = 2*sin(pi * Spearman/6))
  fit.gauss.spearman = tryFitCopula(normalCopula(dim=2,dispstr = "un"), data = U, method = "irho")
  fit.gauss.spearman@loglik = sum( dCopula(U, normalCopula(param=fit.gauss.spearman@estimate,dim=2,dispstr="un"), log=T) )
  # t copula
  fit.t = tryFitCopula(tCopula(dim=2,dispstr="un"), data=U, method="mpl")
  # t copula - method of moments: P^ = sin(pi * Kendall/2)
  fit.t.tau = tryFitCopula(tCopula(dim=2,dispstr="un"), data=U, method="itau.mpl")
  # AC gumbel copula
  fit.gumbel = tryFitCopula(gumbelCopula(dim=2), data=U, method="mpl")
  # AC gumbel copula - method of moments: tau
```

```r
  fit.gumbel.tau = tryFitCopula(gumbelCopula(dim=2), data=U, method="itau")
  fit.gumbel.tau@loglik = sum( dCopula(U, gumbelCopula(param=fit.gumbel.tau@estimate,dim=2), log=T) )
  # AC clayton copula
  fit.clayton = tryFitCopula(claytonCopula(dim=2), data=U, method="mpl")
  # AC clayton copula - method of moments: tau
  fit.clayton.tau = tryFitCopula(claytonCopula(dim=2), data=U, method="itau")
  fit.clayton.tau@loglik = sum( dCopula(U, claytonCopula(param=fit.clayton.tau@estimate,dim=2), log=T) )
  # AC frank copula
  fit.frank = tryFitCopula(frankCopula(dim=2), data=U, method="mpl")
  # AC frank copula - method of moments: tau
  fit.frank.tau = tryFitCopula(frankCopula(dim=2), data=U, method="itau")
  fit.frank.tau@loglik = sum( dCopula(U, frankCopula(param=fit.frank.tau@estimate,dim=2), log=T) )
  res = list(gauss=fit.gauss, gauss.rho=fit.gauss.spearman
             , t=fit.t, t.tau=fit.t.tau
             , frank=fit.frank, frank.tau=fit.frank.tau
             , gumbel=fit.gumbel, gumbel.tau=fit.gumbel.tau
             , clayton=fit.clayton, clayton.tau=fit.clayton.tau)
}
# try to fitCopula or return NA
tryFitCopula = function(...){
  tryCatch(fitCopula(...), error=function(err) NA)
}
# function to obtain p.value of gofCopula of fit using data U
gof.apply = function(fit, U){
  copula = tryCatch(fit@copula, error=function(err) NA)
  # for t-copula, df is fixed and round to the nearest integer
  if (class(copula)=="tCopula") {
    copula@df.fixed=TRUE
    copula@parameters[which(copula@param.names=="df")] = as.integer(round( copula@parameters[which(copula@param.names=="df")] ))
    attr(copula@parameters,"fixed")[which(copula@param.names=="df")] = TRUE
  }
  # set different estim.method
  method = tryCatch(fit@method, error=function(err) NA)
  if ( grepl("rho",method) ) { method="irho" }
  else if ( grepl("tau",method) ) { method="itau" }
  else { method="mpl" }
  # 200 bootstrap are used
  tryCatch(gofCopula(copula,U,200,estim.method=method,simulation="mult",verbose=F)$p.value, error=function(err) NA)
}
# function to obtain loglik of fit
loglik.apply = function(fit){ tryCatch(fit@loglik, error=function(err) NA) }
# function to obtain bic of fit
bic.apply = function(fit){ tryCatch(BIC(fit), error=function(err) NA) }
# function to obtain lambda of fit
lambda.apply = function(fit){ tryCatch(lambda(fit@copula), error=function(err) rep(NA,2)) }
# function to obtain rho of fit
rho.apply = function(fit){ tryCatch(rho(fit@copula), error=function(err) NA) }
# function to obtain tau of fit
tau.apply = function(fit){ tryCatch(tau(fit@copula), error=function(err) NA) }


# Fitting Copulas
options(digits = 4)
# GE:ITT
Fit.GE.ITT = fit.copulas( Uret[,c("GE","ITT")] )
( gof.GE.ITT = sapply(Fit.GE.ITT, gof.apply, Uret[,c("GE","ITT")]) )
sapply(Fit.GE.ITT, bic.apply)
sapply(Fit.GE.ITT, lambda.apply)
sapply(Fit.GE.ITT, rho.apply)
# SPWR:FSLR
Fit.SPWR.FSLR = fit.copulas( Uret[,c("SPWR","FSLR")] )
( gof.SPWR.FSLR = sapply(Fit.SPWR.FSLR, gof.apply, Uret[,c("SPWR","FSLR")]) )
sapply(Fit.SPWR.FSLR, bic.apply)
sapply(Fit.SPWR.FSLR, lambda.apply)
sapply(Fit.SPWR.FSLR, rho.apply)
# FSLR:TSLA
Fit.FSLR.TSLA = fit.copulas( Uret[,c("FSLR","TSLA")] )
( gof.FSLR.TSLA = sapply(Fit.FSLR.TSLA, gof.apply, Uret[,c("FSLR","TSLA")]) )
sapply(Fit.FSLR.TSLA, bic.apply)
sapply(Fit.FSLR.TSLA, lambda.apply)
sapply(Fit.FSLR.TSLA, rho.apply)
# PCG:SPWR
Fit.PCG.SPWR = fit.copulas( Uret[,c("PCG","SPWR")] )
( gof.PCG.SPWR = sapply(Fit.PCG.SPWR, gof.apply, Uret[,c("PCG","SPWR")]) )
sapply(Fit.PCG.SPWR, bic.apply)
sapply(Fit.PCG.SPWR, lambda.apply)
sapply(Fit.PCG.SPWR, rho.apply)


# check scatterplots
```

```r
par(mfrow=c(3,4), cex=0.7)
par(mar=c(4.1, 4.1, 1.1, 2.1))
# log returns
plot.default(RET[,c("GE","ITT")], cex=0.8, col=adjustcolor("black",alpha.f=0.5))
plot.default(RET[,c("SPWR","FSLR")], cex=0.8, col=adjustcolor("black",alpha.f=0.5))
plot.default(RET[,c("FSLR","TSLA")], cex=0.8, col=adjustcolor("black",alpha.f=0.5))
plot.default(RET[,c("PCG","SPWR")], cex=0.8, col=adjustcolor("black",alpha.f=0.5))
# pseudo observations
plot.default(Uret[,c("GE","ITT")], cex=0.8, col=adjustcolor("black",alpha.f=0.3))
plot.default(Uret[,c("SPWR","FSLR")], cex=0.8, col=adjustcolor("black",alpha.f=0.3))
plot.default(Uret[,c("FSLR","TSLA")], cex=0.8, col=adjustcolor("black",alpha.f=0.5))
plot.default(Uret[,c("PCG","SPWR")], cex=0.8, col=adjustcolor("black",alpha.f=0.5))
# simulations
set.seed(123)
plot(Fit.GE.ITT$gauss.rho@copula, n=2012
     , cex=0.8, col=adjustcolor("black",alpha.f=0.3),main="", xlab="gauss.rho U1",ylab="gauss.rho U2")
plot(Fit.SPWR.FSLR$t@copula, n=2012
     , cex=0.8, col=adjustcolor("black",alpha.f=0.3),main="", xlab="t U1",ylab="t U2")
plot(Fit.FSLR.TSLA$frank@copula, n=2012
     , cex=0.8, col=adjustcolor("black",alpha.f=0.5),main="", xlab="frank U1",ylab="frank U2")
plot(Fit.GE.ITT$gauss@copula, n=2012
     , cex=0.8, col=adjustcolor("black",alpha.f=0.5),main="", xlab="gauss U1",ylab="gauss U2")



### (e) PCA ############################################################################

## (e.1) PCA and Construct New Index -----------------------------------

pca = princomp(RET, cor = TRUE)
# percentage of variance explained
pca$sdev^2/sum(pca$sdev^2)
# biplot
fviz_pca_biplot(pca, repel = TRUE, geom = "point", col.var = "black", col.ind = "grey" )

# construct new index and pseudo observations
NewIndex = RET %*% pca$loadings[,1:2]
NewIndex = cbind(NewIndex, SP500)
Uind = apply(NewIndex, 2, edf, adjust=1)
# check sample spearman's rho
cor(Uind, method="spearman")
cor(Uind, method="kendall")
# scatterplot of X and U
pairs2(NewIndex, cex=0.1, col=adjustcolor("black",alpha.f=0.3))
pairs2(Uind, cex=0.1, col=adjustcolor("black",alpha.f=0.3))
# plot of cumulative returns to compare trends
par(mfrow=c(1,1))
cumret = cumprod(exp(NewIndex))-1
cumret = scale(cumret, center=c(0,0,0), scale=c(1,1,1/8))
plot(cumret, legend.loc="topleft", col=c("red","blue","black"), main="Scaled Cumulative Returns")


## (e.2) Fit Copulas -------------------------------------------------------
# Copula for Comp.1 and sp500
Fit.c1 = fit.copulas( Uind[,c(1,3)] )
# check gof, bic, dependence measures
( gof.c1 = sapply(Fit.c1, gof.apply, U=Uind[,c(1,3)]) )
sapply(Fit.c1, bic.apply)
sapply(Fit.c1, lambda.apply)
sapply(Fit.c1, rho.apply)
sapply(Fit.c1, tau.apply)

# Copula for Comp.2 and sp500
Uc2rot = Uind[,c(2,3)]
plot(Uc2rot) # negatively correlated
Uc2rot[,1] = 1-Uc2rot[,1] # rotate s.t. able to fit Gumbel.etc
plot(Uc2rot) # now positively correlated
Fit.c2 = fit.copulas( Uc2rot )
# check gof, bic, dependence measures
( gof.c2 = sapply(Fit.c2, gof.apply, U=Uc2rot) )
sapply(Fit.c2, bic.apply)
sapply(Fit.c2, lambda.apply)
sapply(Fit.c2, rho.apply)
sapply(Fit.c2, tau.apply)

# plot log returns, pseudo observations and simulations
par(mfrow=c(2,3), cex=0.9)
set.seed(2)
plot(as.matrix(NewIndex[,c(1,3)]), col=adjustcolor("black",alpha.f=0.2))
plot(Uind[,c(1,3)], col=adjustcolor("black",alpha.f=0.2))
plot(rCopula(Fit.c1$gauss.rho@copula, n=2012), col=adjustcolor("black",alpha.f=0.2), xlab="gauss.rho U1", ylab="gauss.rho U2")
```

```r
set.seed(2)
tmp = rCopula(Fit.c2$frank.tau@copula, n=2012)
tmp[,1] = 1-tmp[,1]
plot(as.matrix(NewIndex[,c(2,3)]), col=adjustcolor("black",alpha.f=0.5))
plot(Uind[,c(2,3)], col=adjustcolor("black",alpha.f=0.5))
plot(tmp, col=adjustcolor("black",alpha.f=0.5), xlab="frank.tau U1", ylab="frank.tau U2")



### (f) Marshall-Olkin Copula ---------------------------------------------------------------------------

# scatter plot of MO copula simulations
par(mfrow=c(1,3))
plot(moCopula(c(.0, .0)), n = 10000, xaxs="i", yaxs="i", pch = 16, col = adjustcolor("black", 0.2))
plot(moCopula(c(.9, .2)), n = 10000, xaxs="i", yaxs="i", pch = 16, col = adjustcolor("black", 0.2))
plot(moCopula(c(1, 1)), n = 10000, xaxs="i", yaxs="i", pch = 16, col = adjustcolor("black", 0.2))

# scatter plot of exponential transformation
par(mfrow=c(1,4))
tmp = mvrnorm(1000, mu=c(0,0), Sigma=matrix(c(1,0.6,0.6,1),nrow=2))
plot(tmp, xlab="X", ylab="Y")
plot(exp(tmp), xlab="exp(X)", ylab="exp(Y)")
plot(pobs(tmp,ties.method = "max"), xlab="F(X)", ylab="F(Y)")
plot(pobs(exp(tmp),ties.method = "max"), xlab="F(exp(X))", ylab="F(exp(Y))")

#' @title function for computing copula density for MO
#' @param u matrix, pseudo observations
#' @param par vector of length 2, alpha and beta
#' @param log logical, whether to compute log-likelihood
#' @return a vector of copula density
dcopula.MO = function (u, par, log = TRUE) {
  if ( length(par)!=2 ) stop("need 2 params")
  if ( ncol(u)!=2 ) stop("Marshall-Olkin copulas only available in the bivariate case")
  a = par[1]; b = par[2]
  res = apply(u, 1, function(row){
    u = row[1]; v = row[2]
    if (u^a > v^b) { log(1-a)-a*log(u) } else { log(1-b)-b*log(v) }
  })
  if (!log) { res = exp(res) }
  return(res)
}
#' @title function for fitting MO copula using pseudo MLE
#' @param data matrix, original data
#' @param initial vector of length 2, intial values for alpha and beta
#' @return fit information
fit.MO = function (data, initial = c(.5,.5), ...) {
  Udata = pobs(data, ties.method = "max")
  fn = function(par) { -sum( dcopula.MO(Udata, par, log = TRUE) ) }
  fit = optim(initial, fn=fn, upper=c(1-1e-12,1-1e-12), method="L-BFGS-B",...) #, lower=c(0,0), upper=c(1,1)
  return(fit)
}

fit.MO( exp(as.matrix(RET[,c("GE","ITT")])) )
fit.MO( exp(as.matrix(RET[,c("SPWR","FSLR")])) )
fit.MO( exp(as.matrix(RET[,c("FSLR","TSLA")])) )
fit.MO( exp(as.matrix(RET[,c("PCG","SPWR")])) )
fit.MO( exp(as.matrix(NewIndex[,c("Comp.1","SP500")])) )
fit.MO( exp(as.matrix(NewIndex[,c("Comp.2","SP500")])) )
```