Master's Thesis of Norman Doret

Force-directed graph for portfolio selection

-Graph-Oriented Diversification Solution -

포트폴리오 선택을 위한 힘 중심 그래프

August 2022

Graduate School of Engineering
Seoul National University
Computer Science Major

Norman Doret

Force-directed graph for portfolio selection

- Graph-Oriented Diversification Solution -

문병로 교수님

Submitting a master's thesis of
Computer Science

August 2022

Graduate School of Engineering
Seoul National University
Computer Science Major

Norman Doret

Confirming the master's thesis written by
Norman Doret
August 2022

Chair _____(Seal)
Vice Chair _____(Seal)
Examiner _____(Seal)

Abstract

Portfolio diversification is a major concern for a robust investment strategy and time series comparison is maybe the most common way to assess correlation between assets during capital allocation. By creating a graph or network with assets as nodes and pairwise correlation between assets as edges weights, it is possible to identify clusters of assets strongly correlated to the overall market, hence creating a resilient portfolio. Unfortunately, as in many real-world systems, the usual approach for community detection based on shortest path does not account for the real-world conditions.

This research tries to offer constructive insights on the graph building and the correlation computation methods necessary for a good portfolio allocation based on an assets correlation network. This is done through the combination of two research areas: 1. Communicability & centrality measure in graphs and 2. Lower tail dependence for assets correlation assessment.

The final product of this research is a system that takes assets daily return time series as input and output the composition of a portfolio built using an asset correlation network.

Keyword : Portfolio Optimization, Graph, Lower-tail Dependence, Network, Communicability
Student Number : 2020-23584

Table of Contents

Chapter 1. Introduction 1

Chapter 2. Communicability & Centrality 3

Chapter 3. Lowertail Dependence 7

Chapter 4. Computation improvement 9

Chapter 5. Experiments 11

Chapter 6. Results 12

Chapter 7. Discussion 33

Chapter 8. Conclusion 34

Figures 35

Bibliography 40

Abstract in Korean 41

Chapter 1. Introduction

1.1. Study Background

An asset correlation network is a network where nodes represent assets and edges are weighted based on the established correlation of the two assets (nodes) linked by the edge.

Community identification on a network can be done through several methods. Whether it is through edges betweenness computation or hierarchical clustering, they often rely on shortest path or pairwise comparison and rarely account for real-world "side" interactions, where information flow does not necessarily take the shortest path.

To address this issue, Michelle Girvan & M. E. J. Newman studied extensively the notion of communicability [1]: Rather than limiting the interactions between nodes to the shortest path, all paths are considered with a weight inversely scaled to their length. This notion can be extended to other networks' metrics such as centrality [2], which assess how much a node is strongly linked into the overall network①.

While there is evidence of this graph-based strategy being used in the industry②, and while the allocation method once the centrality is computed has been studied, there is little research on which method to use for correlation computation as most graph-based strategies rely on simple correlations such as Pearson correlation or Distance correlation.

① See Betweenness centrality in chapter 2.
② Notably, a portfolio allocation method called Hedgecraft

1.2. Purpose of Research

This paper aims to explore the asset correlation network approach for portfolio diversification while improving the early steps of the process. To establish how much the communicability provides a better representation of the market than the simple edge betweenness, and by using different correlation assessments, a comparison is made of this network-based method on pools of stocks from well-known indexes.

While distance correlation (used in the initial strategy) takes non-linear correlation into account, it has a limited interpretation in the real world. Another common method in finance to assess the correlation between two asset is the lower tail dependence. Applied on stock returns, it describes how much an asset price would be impacted knowing that an other's is going to zero. This is especially useful to provide robust strategy in trying times such as crisis or high volatility periods.

Based on the work of Giovanni De Luca and Paola Zuccolotto [6], an attempt is made to build the market graph with lower tail dependence as the correlation value. It is then compared to the initial approach using distance correlation, as well as approaches only based on shortest path methods for the assets' allocation. Every experiment is conducted on S&P500, DAX, CAC and Kospi underlying assets to test the robustness of the strategy using different experiment parameters and on different markets.

Chapter 2. Communicability & Centrality

2.1. Communicability

The first limit to network representation of real-world complex systems is how the interactions between nodes are evaluated. Common methods rely on shortest path while reality is not as simple. Even though it is more obvious for mechanical or thermodynamical systems, finance is not spared by this inherent complexity. Hence, it is necessary to consider other paths than the shortest one in an asset correlation network.

To that extent, we need the followings:

Given an unoriented graph G such as
$G = (V, E),\ |V| = n,\ |E| = m,\ A(G) = A \in \{0,1\}^{n \times n}$

We define the following values:

number of walks of length k from node p to q	number of shortest paths of length s between nodes p and q	the number of walks of length $k > s$ connecting the nodes p and q
$(A^k)_{p,q}$	$P_{pq}^{(s)}$	$W_{pq}^{(k)}$

Using those, we can define the communicability between p and q as

$$G_{pq} = \frac{1}{s!}P_{pq}^{(s)} + \sum_{k>s}\frac{1}{k!}W_{pq}^{(k)} = \sum_{j=1}^{\infty}\frac{(A^k)_{pq}}{k!} = (e^A)_{pq}$$

Rather than considering solely the shortest path between p and q, this communicability takes all paths into account, inversely weighted by their length. Moreover, as the definition allows it, the communicability is none other than the exponential of the adjacency matrix. This property is extremely convenient since given $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ the eigenvalues of A and $\varphi_j(p)$ the p^{th} element of

the j^{th} orthonormal eigenvector of the adjacency matrix (associated with λ_j), we can express the communicability between p and q as

$$G_{pq} = \sum_{j=1}^{\infty} \frac{(A^k)_{pq}}{k!} = e^A = \sum_{j=1}^{n} \varphi_j(p)\varphi_j(q)e^{\lambda_j}$$

2.2. Green function

Green's function is defined as the impulse response of an inhomogeneous linear differential operator in a domain with specified initial conditions. Estrada & Hatano [5] shows that the communicability can be expressed as the Green's function of the network. By treating each node as an oscillator and each edge as a spring, we can express the mechanical system that ensue as follow:

Let F_p be the force from p applying on q (1), we can then derivate the potential energy from the resulting force (2). By summing these, we find the total energy (3). Diagonalizing L we can express the partition function (4) as in (5)

(1) $F_p = K\sum_q (z_p - z_q)A_{pq}$ (K a common spring constant)

(2) $U_p = \frac{K}{2}\sum_q (z_p - z_q)^2 A_{pq}$

(3) $E = \sum_p U_p = \frac{K}{2}\sum_{p,q}(z_p - z_q)^2 A_{pq} = -K\sum_{p,q} z_p L_{pq} z_q$

$L_{pq} = A_{pq} - k_p\delta_{pq}$ (L being the Laplacian matrix of the graph)

(4) $Z = \sum_{all\ config} e^{-\beta E} = \frac{1}{Z}\int z_p z_q \exp(\beta K \sum_{s,l} z_s L_{sl} z_l) \prod_r dz_r$

(5) $$Z = \frac{1}{Z}\int \exp(\beta K \sum_{s,l} \lambda_j u_j^2) \prod_j du_j$$

$$G_{pq}(\beta) = \langle z_p | z_q \rangle = \frac{1}{Z} \int z_p z_q \exp\left(\beta K \sum_{s,l} z_s L_{sl} z_l\right) \prod_r dz_r$$

G_{pq} is the Green's function of the network and represents how much node q oscillate when node p is shaken. It is even possible to extend that definition to weighted graphs as the symmetric matrix still allows diagonalization:

$$F_p = K \sum_q (z_p - z_q) A_{pq} = \sum_q k_{pq} (z_p - z_q)$$

The asset correlation network can then be represented as a force-directed graph where each edge is a spring subject to the Hooke law. When an information (an impulse) is released, the market is impacted, and the contagion spread from assets to assets. The assets prices are impacted further propagate the phenomenon.

2.3. Centrality

Unfortunately, identifying communities using solely the communicability is difficult for asset correlation networks. In their work, Michelle Girvan & M. E. J. Newman [2] identify two main graph structures with specific communicability properties:

- Disassortative: strong communicability between hubs and nodes of low degree
- Assortative: strong communicability between nodes with the highest degrees (hubs)

The main drawback to asset correlation networks is that they are of the latter kind and tend to show overlapping communities, making it difficult to properly distribute the capital among them. Instead, a solution is to consider the centrality of a node to the network. Rather than allocating capital among communities of assets, we allocate the capital based on how correlated the assets are to the overall market.

The common centrality of node v is based on shortest path as

follow:

$$c_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

But we want to use the communicability previously established to generalize the centrality:

$$c_{CB}(v) = \frac{1}{C} \sum_{s \neq v \neq t} \frac{G_{st}(v)}{G_{st}}$$

A representation of the asset correlation network for the German (DAX) and French (CAC) indexes' stocks, with assets colored by centrality can be observed in Figure 1. The corresponding asset allocation is then visible in the Results section.

In order to amplify the difference of centrality between assets and to reduce the computation cost, it is possible to prune away the edges with smallest weights. Estrada, Higham & Hatano [5] recommend pruning edges with weights below 0.325 but in order to preserve the connexe structure of the graph, it is sometimes necessary to use a lower threshold (0.2 for DAX for example).

Chapter 3. Lower tail Dependence

3.1. Lower tail dependence

While distance correlation is commonly used to assess correlation between time series, a lot of other methods are used in finance. In order to improve the results of the performance of this network-based strategy, we need to find a correlation computation that better fit the assumptions made when using Green's function. The correlation coefficient represents the elasticity constant of the edges in the network. Hence it needs to account for the "impulse propagation" mechanic that is represented by the communicability. In finance, this could be explained as the "contagion" phenomenon triggered by bad news. An unexpected event drives the price of an asset down and this news propagate to others similar assets that are then affected (for example, a shortage of semi-conductor might first impact electronics companies before spreading to the overall tech market).

Lower tail dependence assesses how much an asset's price is likely to tend to zero knowing another one tends to zero. It can be formally expressed as following:

$$\lambda_l = \lim_{q \to 0} P(X_2 \leq F_2^{\leftarrow}(q) | X_1 \leq F_1^{\leftarrow}(q))$$

where $F^{\leftarrow}(q) = \inf\{x \in \mathbb{R} : F(x) \geq q\}$

In practice, the tail dependence coefficients must be estimated from observed data. A very effective way of modeling financial returns is to use a copula function thanks to which tail dependence estimation is both simple and flexible.

The difference between distance correlation and lower tail dependence can be observed in Figure 2 using hierarchical clustering.

3.2. Copulas

A copula is a multivariate cumulative distribution function defined as follow: $C:[0,1]^d \rightarrow [0,1]$ (in our case, with two time series, $d = 2$, a bivariate copula). It describes the dependence structure between the variables. The main advantage of copulas resides in the Sklar's theorem that states that every multivariate cumulative distribution function can be expressed in term of its marginals and a copula.

For the lower tail dependence, the most used class of copula is the class of Archimedean copulas as they often admit an explicit formula and allow modeling dependence in high dimensions. Specifically, in our case, we use the Clayton copula, defined as follow:

$$C_\theta(u,v) = \left[\max\left(u^{-\theta} + v^{-\theta} - 1; 0\right)\right]^{-\frac{1}{\theta}} \text{ with } \theta \in \mathbb{R}\backslash\{0\}$$

The process to determine the lower tail dependence is the following: Empirically fit the copula $\theta \in \mathbb{R}\backslash\{0\}$ parameters on the time series studied, derivate marginal distribution from the copula and finally determine lower tail dependence.

Chapter 4. Computation improvement

While the graph building and the centrality computation have a relatively light computational cost, the copula becomes extremely slow to process for a large number of assets (during an S&P500 portfolio construction for example). Moreover, the need for dynamic strategy or the will to apply this strategy on a shorter-term basis may require reducing this computation time.

De Luca & Zuccolotto [7] have shown evidence of association between index volatility and lower tail dependence of pairs of assets. This provides an opportunity to improve the computation time by deducing the copula parameter directly from the index volatility using a simple linear regression.

To assess the potential gain in computational power, the following experiment is realized for CAC and DAX indexes:

- Compute the index daily volatility using the ARCH model
- Compute lower tail dependence on a daily rolling period using a copula for each pair of assets
- Plot the scatter points of the resulting values $(\lambda_{ij,t})_{t\in[0,T]}$ against the index volatility $(\sigma_t)_{t\in[0,T]}$

While the results are not as conclusive as De Luca & Zuccolotto's on the Italian market, Figure 3 shows that some scatter plots still provide evidence of a strong linear correlation. Applying a linear regression to those gives us an estimate value for the copula parameter of the pair.

For each pair with a sufficient linear correlation factor, the following estimation is made to replace the parameter fitting during the lower tail dependence computation: $\theta_{t,ij} = \exp(\omega_{ij} + \alpha_{ij}\sigma_{t-1})$. The lower tail dependence is then estimated as follow:

$$\lambda_{Lt\,ij} \approx 2^{-\frac{1}{\theta_{t,ij}}}$$

De Luca & Zuccolotto note that pairs with significant correlation have a positive α_{ij}. In their own words, “The lower tail dependence coefficient tends to increase with rising volatility in the market, in accordance with the idea of contagion”. This confirms the legitimacy of using the lower tail dependence in our centrality-based approach.

Chapter 5. Experiments

5.1. Strategy

Once each asset is given a centrality value, we allocate a share of capital depending on this value. By allocating capital to assets depending on their centrality (the lower the centrality the higher the share), we create a portfolio disconnected from the global market trend, hence more resilient to high volatility periods such as those seen recently.

$$M = \begin{pmatrix} \lambda_{00} & \cdots & \lambda_{0n} \\ \vdots & \ddots & \vdots \\ \lambda_{n0} & \cdots & \lambda_{nn} \end{pmatrix}, \; centrality\big(G(M)\big) = (c_i)_{i\in[1,n]}$$

Four allocation functions are proposed:

Avg	Exp	Ln	TT
$W_i = \frac{(1-c_i)}{C_{\text{avg}}}$	$W_i = \frac{\exp(-c_i)}{C_{\exp}}$	$W_i = \frac{-\ln(c_i)}{C_{ln}}$	$W_i = \frac{(c_{max}-c_i)}{C_{\text{tt}}}$

With C_{avg} , C_{exp} , C_{ln} , C_{tt} as normalization constants

5.2. Back test

The correlations matrix used as adjacency matrix to build the graphs are created using 2015−2018 daily returns from S&P, Kospi, DAX and CAC underlying assets[③]. For each index, distance correlation and lower−tail dependence matrix are computed. For each correlation matrix, several graphs are created with increasing threshold[④] to prune the edges with smallest correlation values. The nodes centrality of each graph is then computed and turned into an asset allocation using the four allocation functions. All these newly formed portfolios are tested on 2018−2021 market data.

[③] The stocks that were included or excluded from the index during the training period are not considered.

[④] 0.3 for S&P and CAC, 0.2 for DAX, 0.03 for Kospi

Chapter 6. Results

Table 1. Return (in %) on the 2018−2021 period for every strategy.

	Uniform	*Index*
S&P	57.74	39.33
Kospi	19.93	19.41
DAX	26	6.58
CAC	2.09	4.97

Lowertail

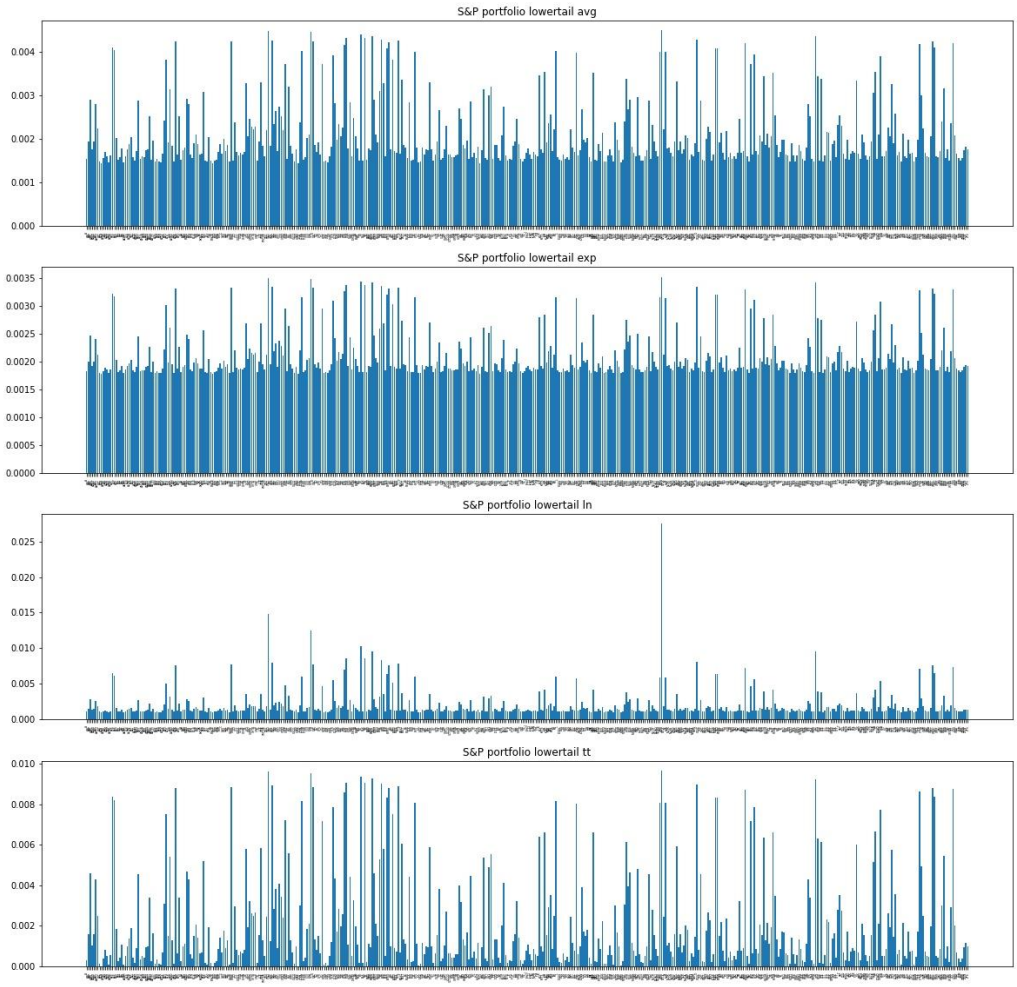
	Communicability				Shortest path			
	Avg	Exp	Ln	TT	Avg	Exp	Ln	TT
S&P	71.66	65.87	107.66	101.89	57.75	57.75	58.63	61.08
Kospi	21.23	20.43	19.96	23.57	19.94	19.94	20.62	22.6
DAX	42.32	34.66	104.31	180.96	25.95	25.95	25.7	24.16
CAC	4.98	3.7	20.13	26.96	2.06	2.06	1.97	0.87

Distance Correlation

	Communicability				Shortest path			
	Avg	Exp	Ln	TT	Avg	Exp	Ln	TT
S&P	54.74	56.55	53.75	21.22	57.74	57.74	56.59	23.11
Kospi	19.93	19.93	19.93	13.29	19.93	19.93	*[5]	*
DAX	25.7	25.89	25.6	−15.33	26	26	25.61	−15.33
CAC	1.62	1.9	1.45	−11.77	2.09%	2.09%	1.6	0.87

[5] Because of the low threshold for edges pruning, these allocation functions could not be computed.

5.1. S&P


S&P portfolio lowertail avg
0.004
0.003
0.002
0.001
0.000
S&P portfolio lowertail exp
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
S&P portfolio lowertail ln
0.025
0.020
0.015
0.010
0.005
0.000
S&P portfolio lowertail tt
0.010
0.008
0.006
0.004
0.002
0.000

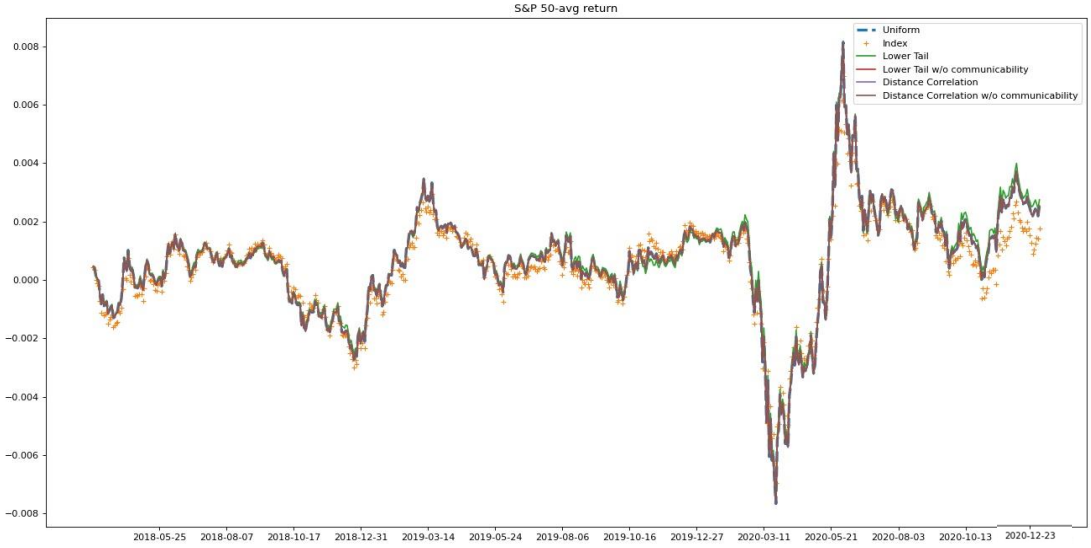
AVG


S&P capital
1e6
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
1.6
1.4
1.2
1.0
0.8
2018-01-02
2018-03-15
2018-05-25
2018-08-07
2018-10-17
2018-12-31
2019-03-14
2019-05-24
2019-08-06
2019-10-16
2019-12-27
2020-03-11
2020-05-21
2020-08-03
2020-10-13
2020-12-23

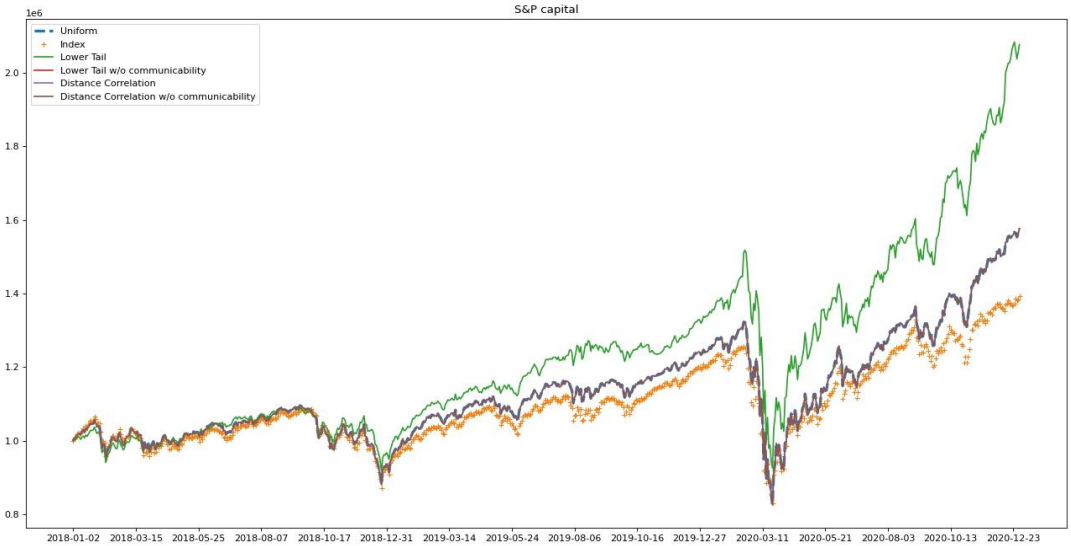

S&P 50-avg return
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
0.008
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.008
2018-05-25
2018-08-07
2018-10-17
2018-12-31
2019-03-14
2019-05-24
2019-08-06
2019-10-16
2019-12-27
2020-03-11
2020-05-21
2020-08-03
2020-10-13
2020-12-23

EXP


S&P capital
1e6
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
1.6
1.4
1.2
1.0
0.8
2018-01-02
2018-03-15
2018-05-25
2018-08-07
2018-10-17
2018-12-31
2019-03-14
2019-05-24
2019-08-06
2019-10-16
2019-12-27
2020-03-11
2020-05-21
2020-08-03
2020-10-13
2020-12-23

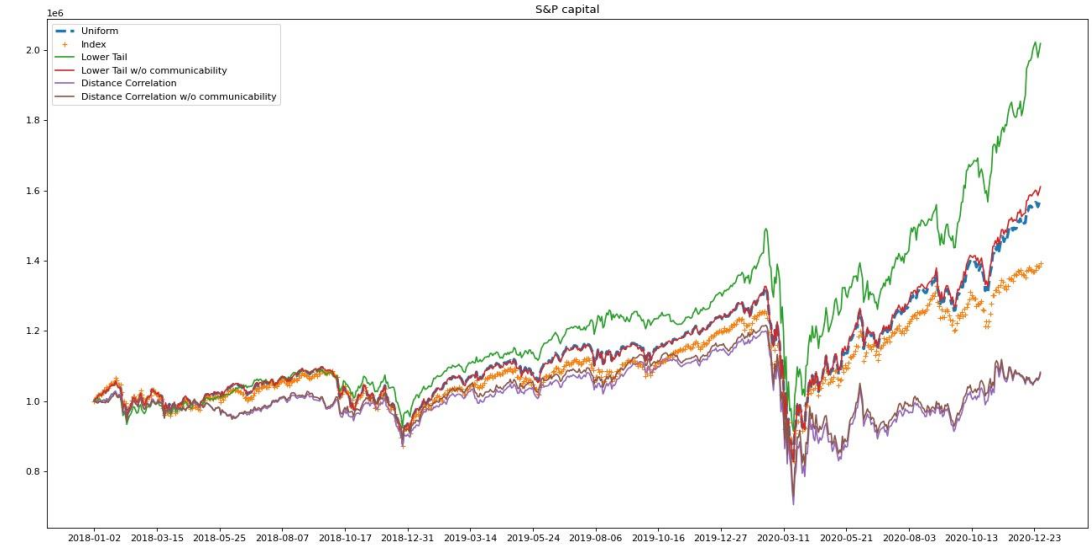

S&P 50-avg return
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
0.008
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.008
2018-05-25
2018-08-07
2018-10-17
2018-12-31
2019-03-14
2019-05-24
2019-08-06
2019-10-16
2019-12-27
2020-03-11
2020-05-21
2020-08-03
2020-10-13
2020-12-23

LN


S&P capital
1e6
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
2.0
1.8
1.6
1.4
1.2
1.0
0.8
2018-01-02
2018-03-15
2018-05-25
2018-08-07
2018-10-17
2018-12-31
2019-03-14
2019-05-24
2019-08-06
2019-10-16
2019-12-27
2020-03-11
2020-05-21
2020-08-03
2020-10-13
2020-12-23


S&P 50-avg return
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
0.008
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.008
2018-05-25
2018-08-07
2018-10-17
2018-12-31
2019-03-14
2019-05-24
2019-08-06
2019-10-16
2019-12-27
2020-03-11
2020-05-21
2020-08-03
2020-10-13
2020-12-23

TT

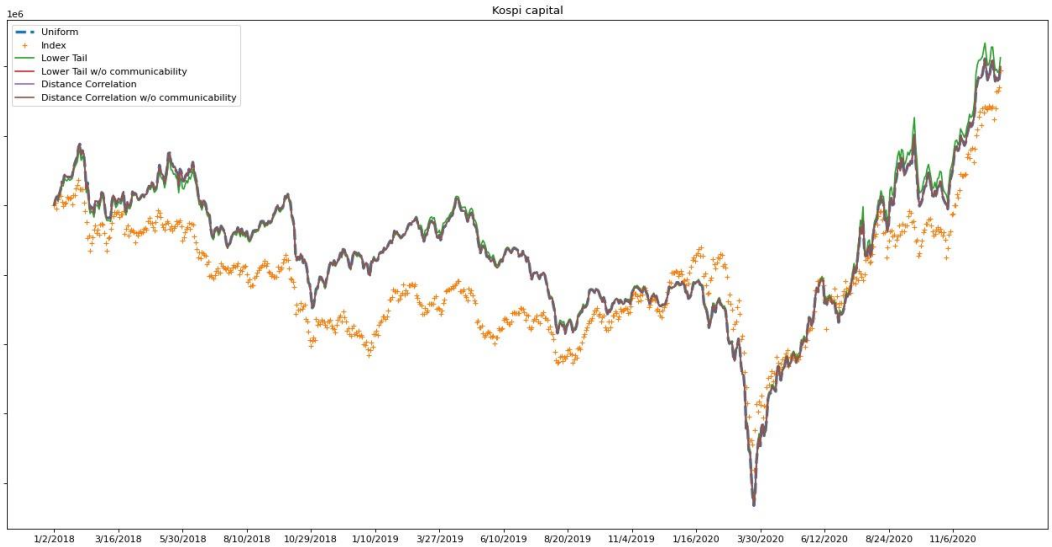

S&P capital
1e6
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
2.0
1.8
1.6
1.4
1.2
1.0
0.8
2018-01-02
2018-03-15
2018-05-25
2018-08-07
2018-10-17
2018-12-31
2019-03-14
2019-05-24
2019-08-06
2019-10-16
2019-12-27
2020-03-11
2020-05-21
2020-08-03
2020-10-13
2020-12-23

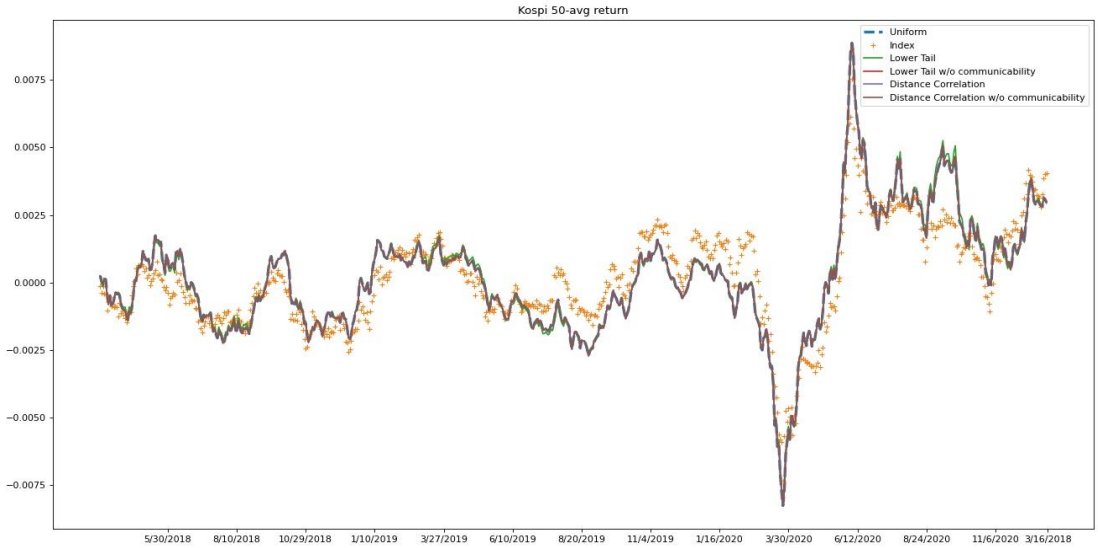

S&P 50-avg return
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
0.0075
0.0050
0.0025
0.0000
−0.0025
−0.0050
−0.0075
2018-05-25
2018-08-07
2018-10-17
2018-12-31
2019-03-14
2019-05-24
2019-08-06
2019-10-16
2019-12-27
2020-03-11
2020-05-21
2020-08-03
2020-10-13
2020-12-23

5.2. Kospi

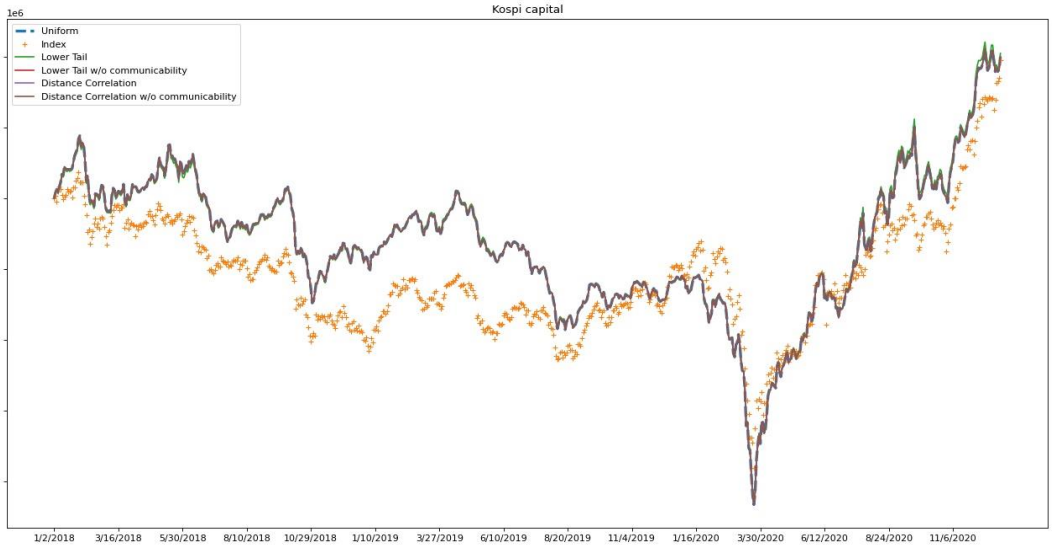

Kospi portfolio lowertail avg
Kospi portfolio lowertail exp
Kospi portfolio lowertail ln
Kospi portfolio lowertail tt

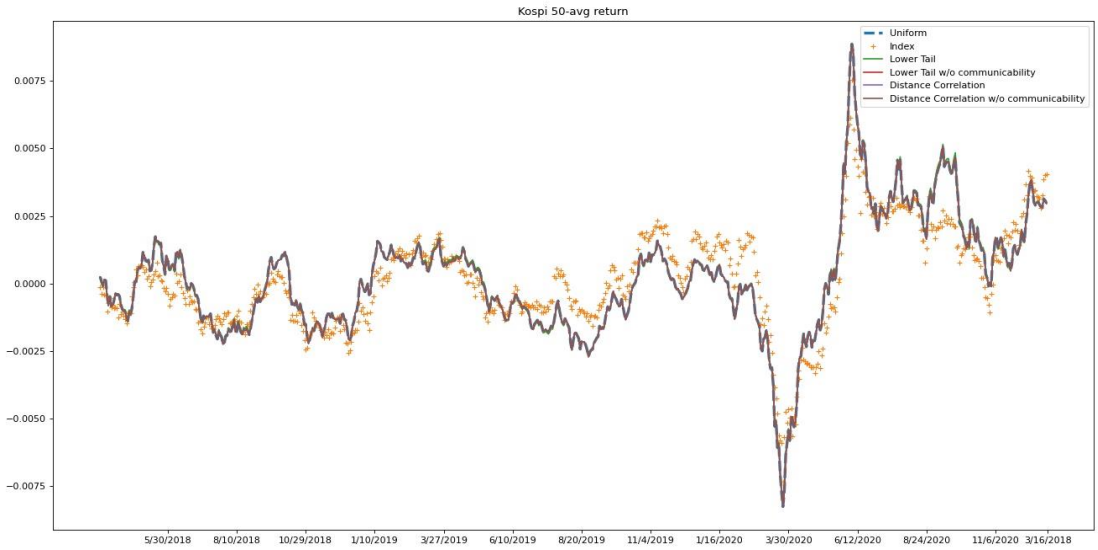
AVG


Kospi capital
1e6
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
1.2
1.1
1.0
0.9
0.8
0.7
0.6
1/2/2018
3/16/2018
5/30/2018
8/10/2018
10/29/2018
1/10/2019
3/27/2019
6/10/2019
8/20/2019
11/4/2019
1/16/2020
3/30/2020
6/12/2020
8/24/2020
11/6/2020

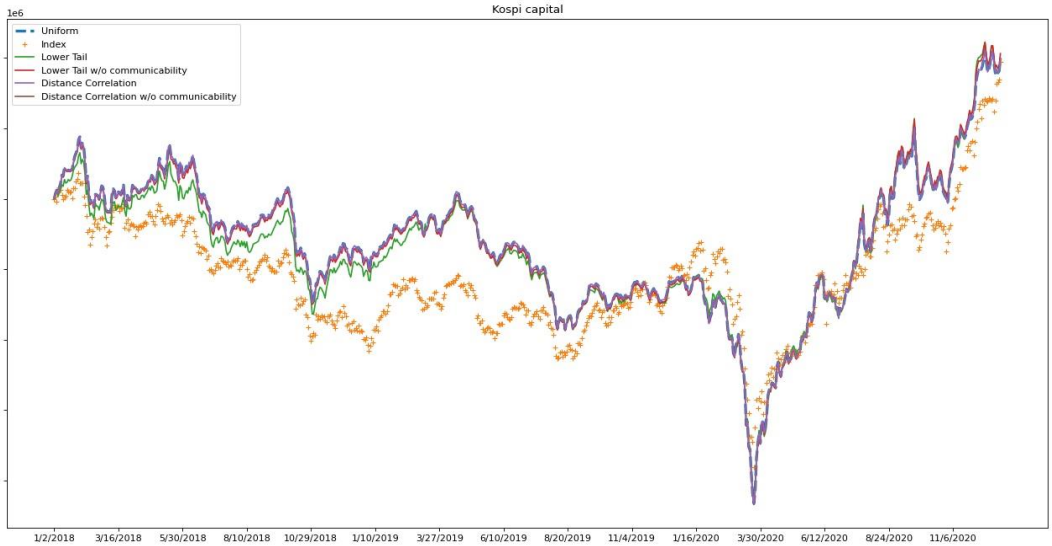

Kospi 50-avg return
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
0.0075
0.0050
0.0025
0.0000
−0.0025
−0.0050
−0.0075
5/30/2018
8/10/2018
10/29/2018
1/10/2019
3/27/2019
6/10/2019
8/20/2019
11/4/2019
1/16/2020
3/30/2020
6/12/2020
8/24/2020
11/6/2020
3/16/2018

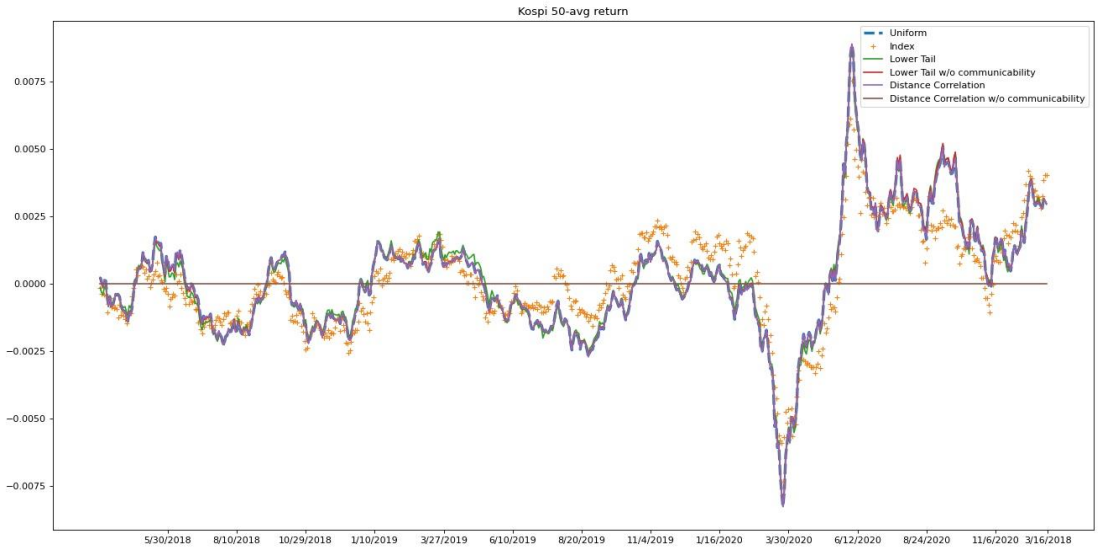
EXP


Kospi capital
1e6
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
1.2
1.1
1.0
0.9
0.8
0.7
0.6
1/2/2018
3/16/2018
5/30/2018
8/10/2018
10/29/2018
1/10/2019
3/27/2019
6/10/2019
8/20/2019
11/4/2019
1/16/2020
3/30/2020
6/12/2020
8/24/2020
11/6/2020

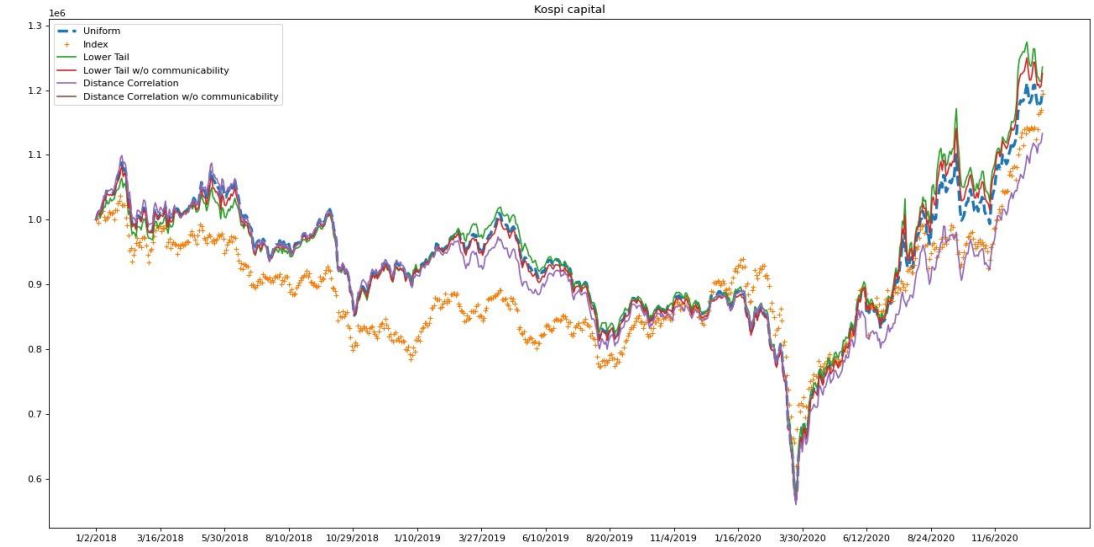

Kospi 50-avg return
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
0.0075
0.0050
0.0025
0.0000
−0.0025
−0.0050
−0.0075
5/30/2018
8/10/2018
10/29/2018
1/10/2019
3/27/2019
6/10/2019
8/20/2019
11/4/2019
1/16/2020
3/30/2020
6/12/2020
8/24/2020
11/6/2020
3/16/2018

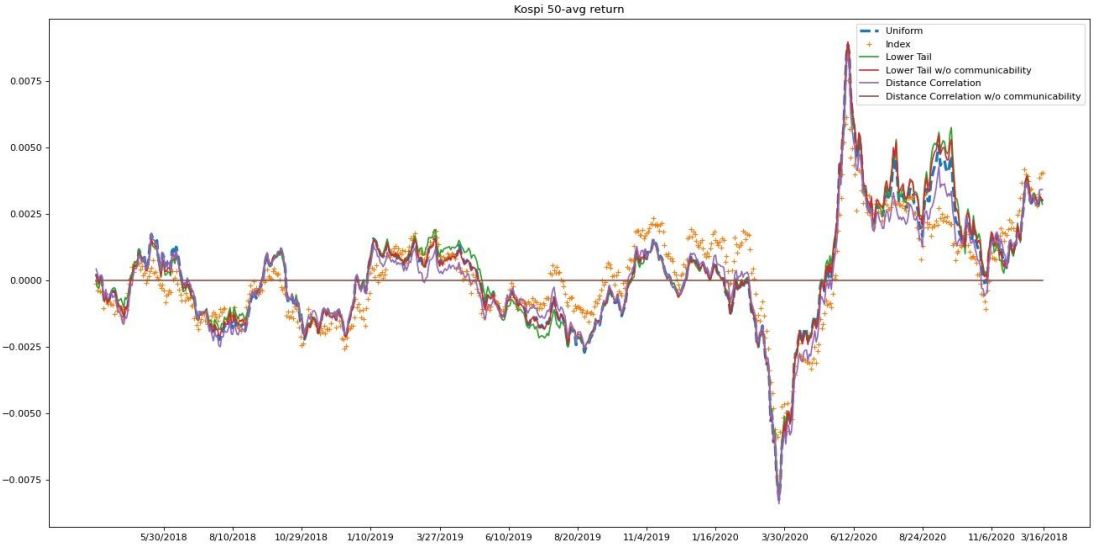
LN


Kospi capital
1e6
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
1.2
1.1
1.0
0.9
0.8
0.7
0.6
1/2/2018
3/16/2018
5/30/2018
8/10/2018
10/29/2018
1/10/2019
3/27/2019
6/10/2019
8/20/2019
11/4/2019
1/16/2020
3/30/2020
6/12/2020
8/24/2020
11/6/2020

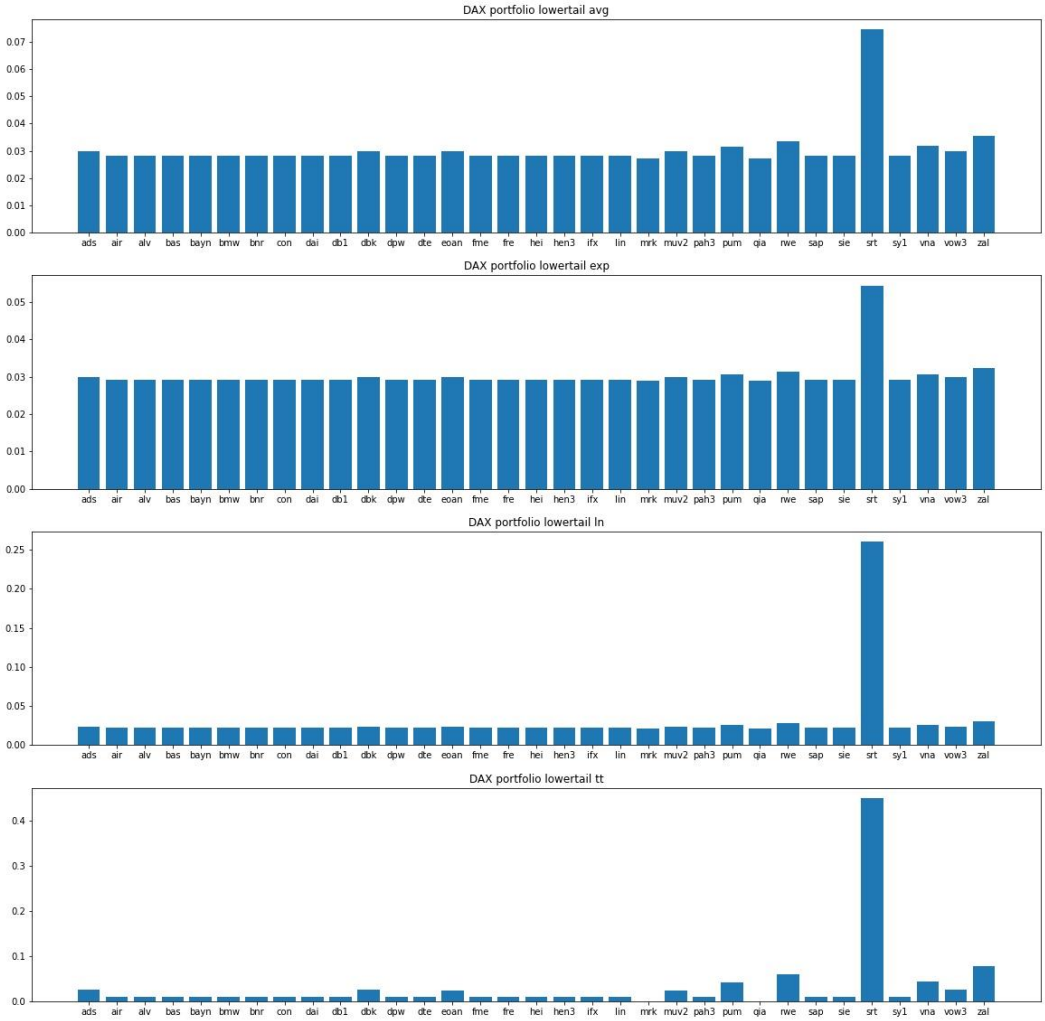

Kospi 50-avg return
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
0.0075
0.0050
0.0025
0.0000
−0.0025
−0.0050
−0.0075
5/30/2018
8/10/2018
10/29/2018
1/10/2019
3/27/2019
6/10/2019
8/20/2019
11/4/2019
1/16/2020
3/30/2020
6/12/2020
8/24/2020
11/6/2020
3/16/2018

TT

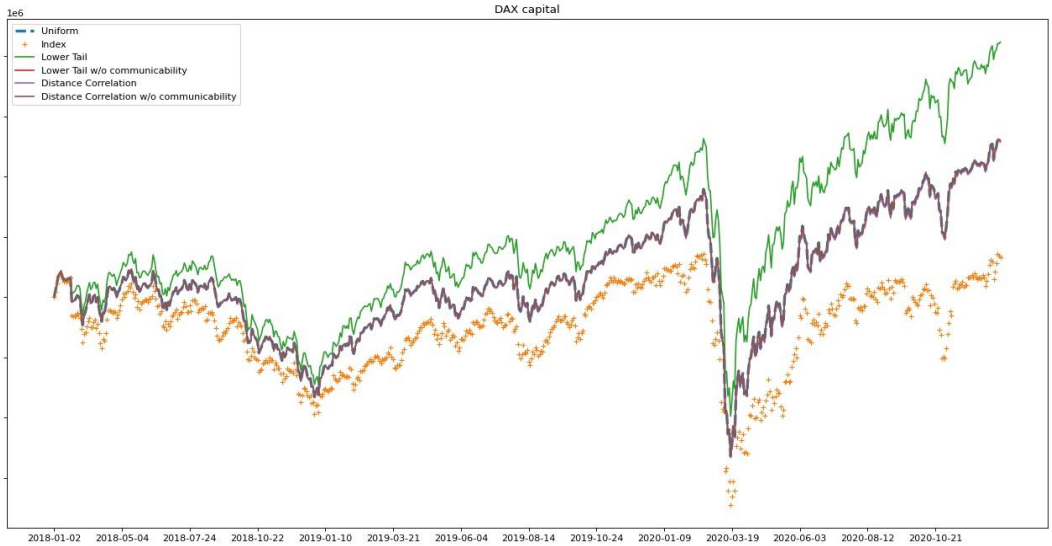

Kospi capital
1e6
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
1/2/2018
3/16/2018
5/30/2018
8/10/2018
10/29/2018
1/10/2019
3/27/2019
6/10/2019
8/20/2019
11/4/2019
1/16/2020
3/30/2020
6/12/2020
8/24/2020
11/6/2020


Kospi 50-avg return
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
0.0075
0.0050
0.0025
0.0000
−0.0025
−0.0050
−0.0075
5/30/2018
8/10/2018
10/29/2018
1/10/2019
3/27/2019
6/10/2019
8/20/2019
11/4/2019
1/16/2020
3/30/2020
6/12/2020
8/24/2020
11/6/2020
3/16/2018

5.3. DAX

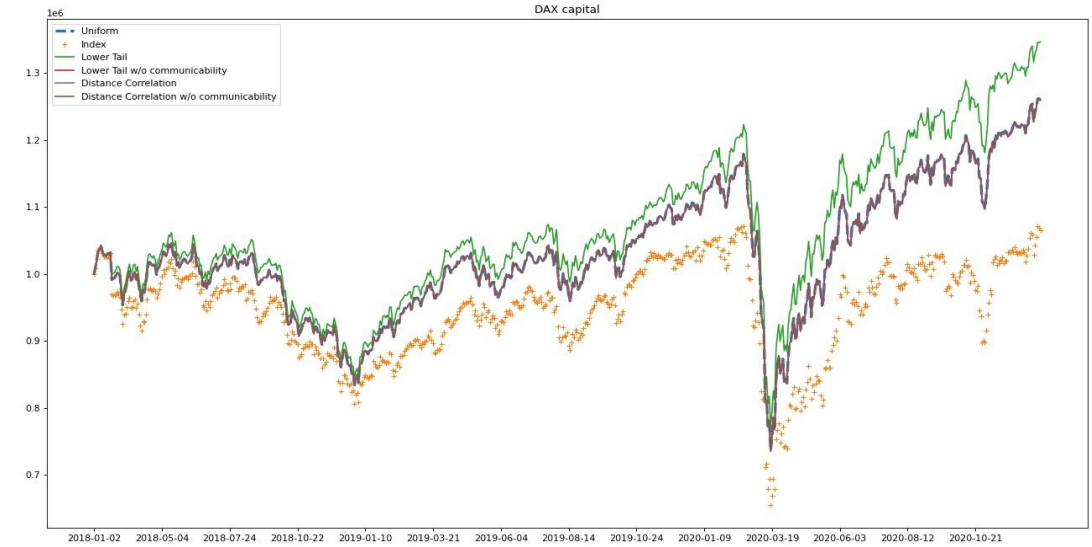

DAX portfolio lowertail avg
DAX portfolio lowertail exp
DAX portfolio lowertail ln
DAX portfolio lowertail tt
ads air alv bas bayn bmw bnr con dai db1 dbk dpw dte eoan fme fre hei hen3 ifx lin mrk muv2 pah3 pum qia rwe sap sie srt sy1 vna vow3 zal

AVG


DAX capital
1e6
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
2018-01-02
2018-05-04
2018-07-24
2018-10-22
2019-01-10
2019-03-21
2019-06-04
2019-08-14
2019-10-24
2020-01-09
2020-03-19
2020-06-03
2020-08-12
2020-10-21

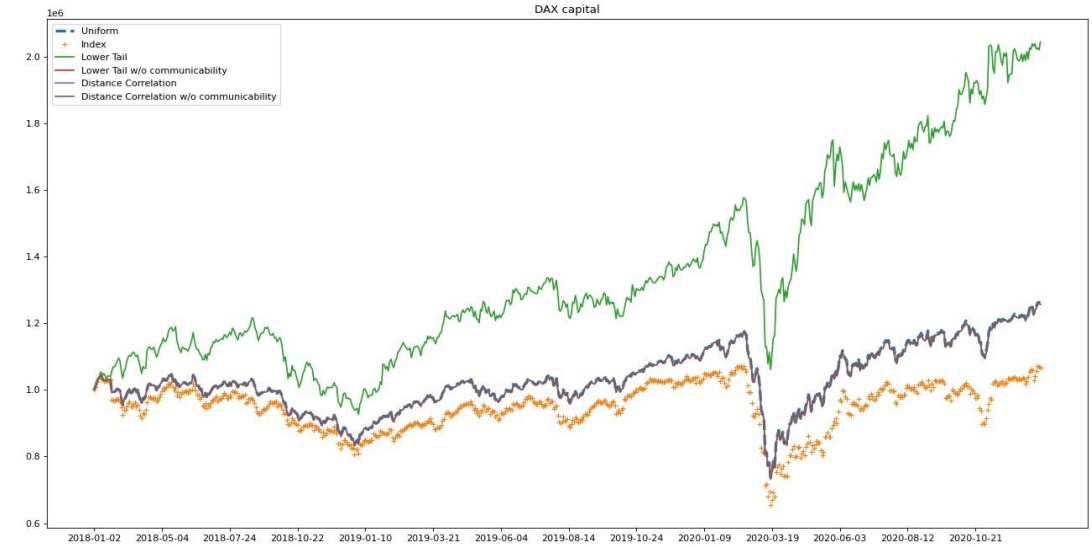

DAX 50-avg return
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
0.0075
0.0050
0.0025
0.0000
−0.0025
−0.0050
−0.0075
2018-07-24
2018-10-22
2019-01-10
2019-03-21
2019-06-04
2019-08-14
2019-10-24
2020-01-09
2020-03-19
2020-06-03
2020-08-12
2020-10-21
2018-05-04

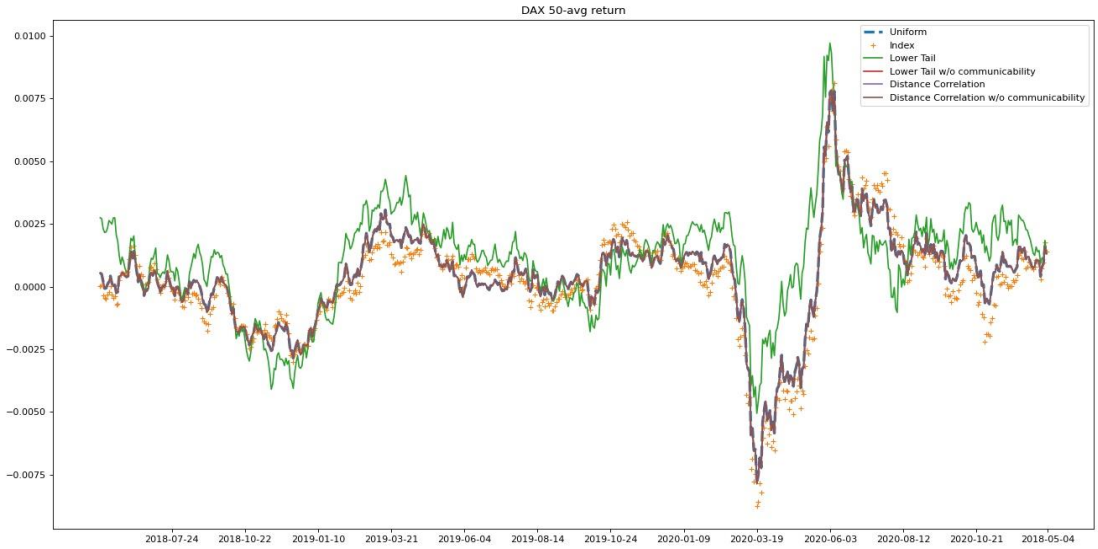
EXP


DAX capital
1e6
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
1.3
1.2
1.1
1.0
0.9
0.8
0.7
2018-01-02
2018-05-04
2018-07-24
2018-10-22
2019-01-10
2019-03-21
2019-06-04
2019-08-14
2019-10-24
2020-01-09
2020-03-19
2020-06-03
2020-08-12
2020-10-21

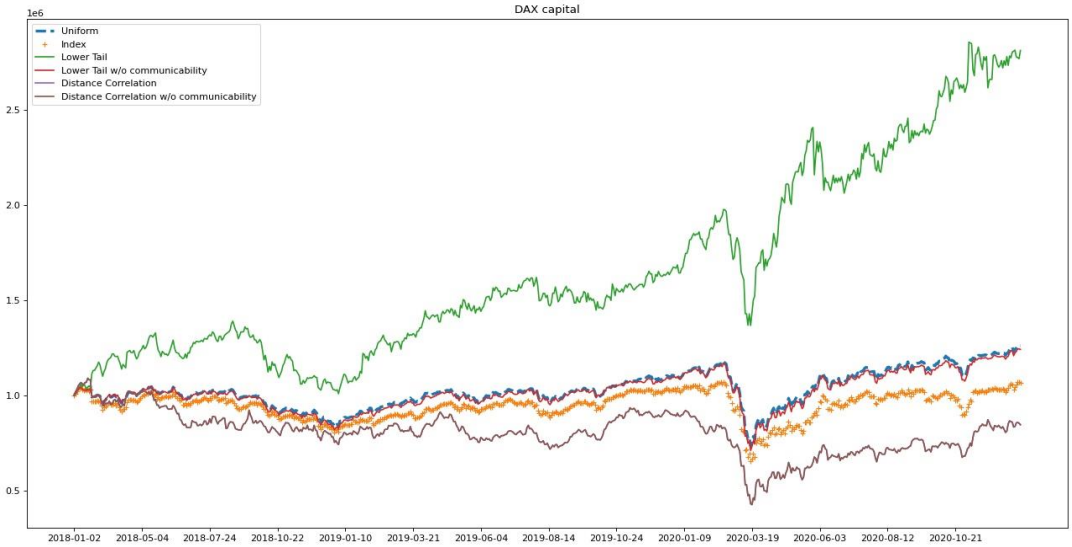

DAX 50-avg return
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
0.0075
0.0050
0.0025
0.0000
−0.0025
−0.0050
−0.0075
2018-07-24
2018-10-22
2019-01-10
2019-03-21
2019-06-04
2019-08-14
2019-10-24
2020-01-09
2020-03-19
2020-06-03
2020-08-12
2020-10-21
2018-05-04

LN


DAX capital
1e6
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
2018-01-02
2018-05-04
2018-07-24
2018-10-22
2019-01-10
2019-03-21
2019-06-04
2019-08-14
2019-10-24
2020-01-09
2020-03-19
2020-06-03
2020-08-12
2020-10-21

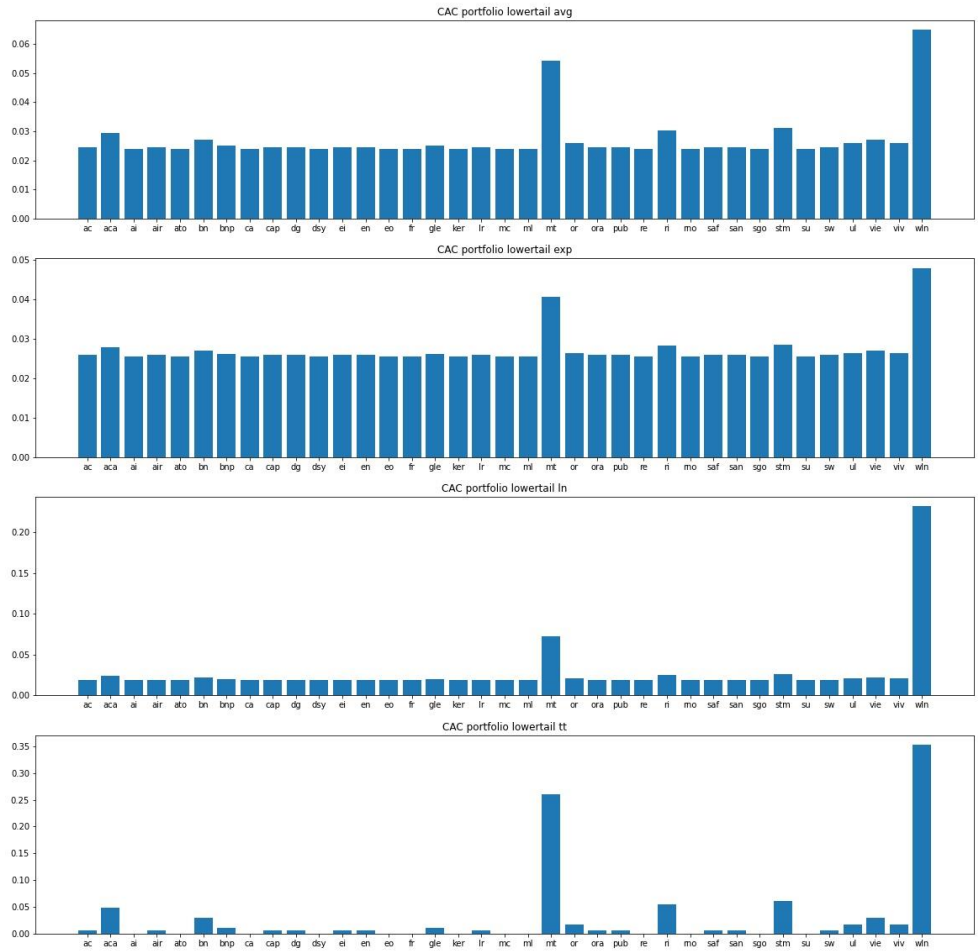

DAX 50-avg return
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
0.0100
0.0075
0.0050
0.0025
0.0000
−0.0025
−0.0050
−0.0075
2018-07-24
2018-10-22
2019-01-10
2019-03-21
2019-06-04
2019-08-14
2019-10-24
2020-01-09
2020-03-19
2020-06-03
2020-08-12
2020-10-21
2018-05-04

TT

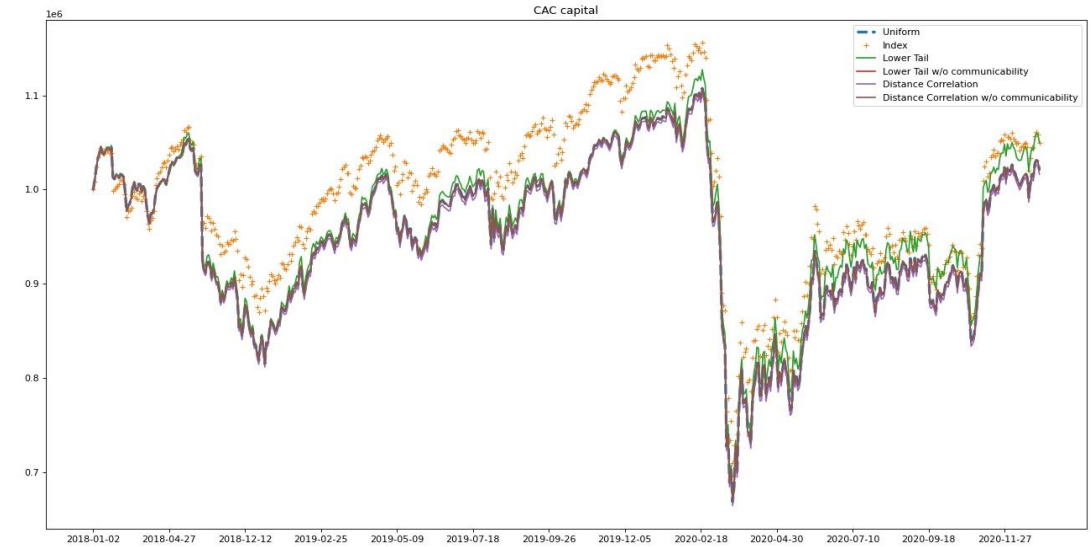

DAX capital
1e6
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
2.5
2.0
1.5
1.0
0.5
2018-01-02
2018-05-04
2018-07-24
2018-10-22
2019-01-10
2019-03-21
2019-06-04
2019-08-14
2019-10-24
2020-01-09
2020-03-19
2020-06-03
2020-08-12
2020-10-21

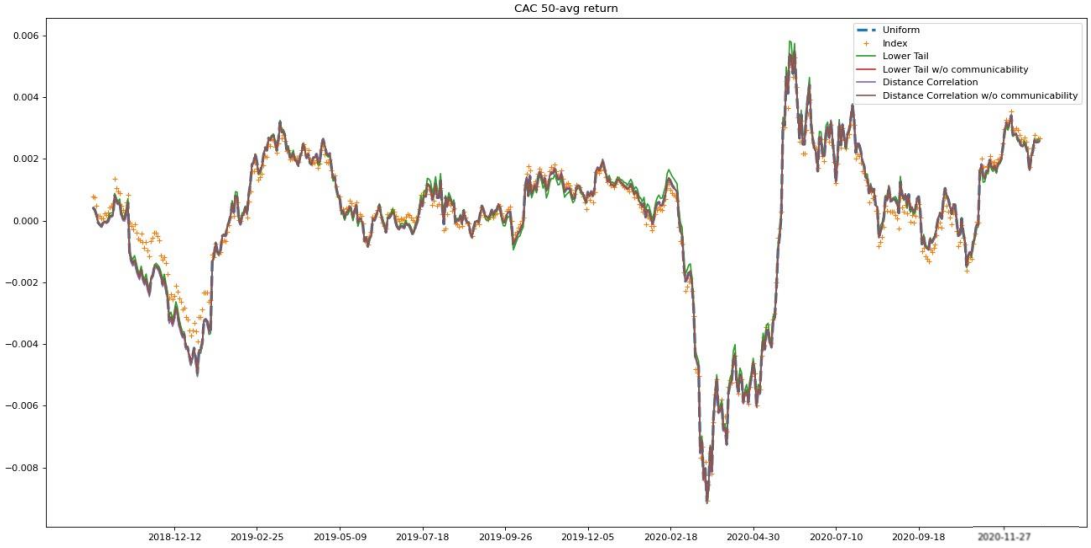

DAX 50-avg return
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
0.010
0.005
0.000
−0.005
−0.010
−0.015
2018-07-24
2018-10-22
2019-01-10
2019-03-21
2019-06-04
2019-08-14
2019-10-24
2020-01-09
2020-03-19
2020-06-03
2020-08-12
2020-10-21
2018-05-04

5.4. CAC

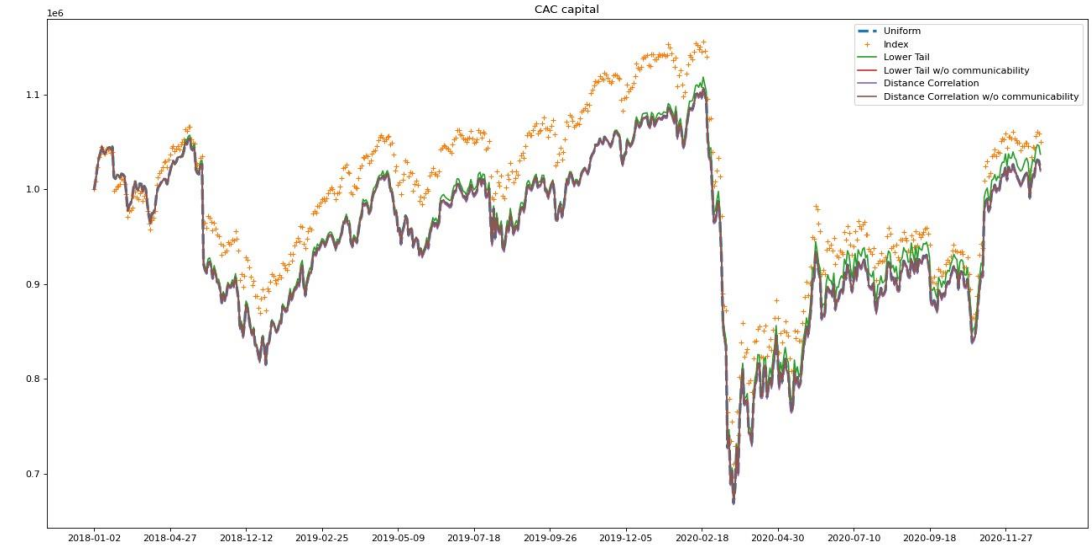

CAC portfolio lowertail avg
CAC portfolio lowertail exp
CAC portfolio lowertail ln
CAC portfolio lowertail tt
ac aca ai air ato bn bnp ca cap dg dsy ei en eo fr gle ker lr mc mil mit or ora pub re ri mo saf san sgo stm su sw ul vie viv wln

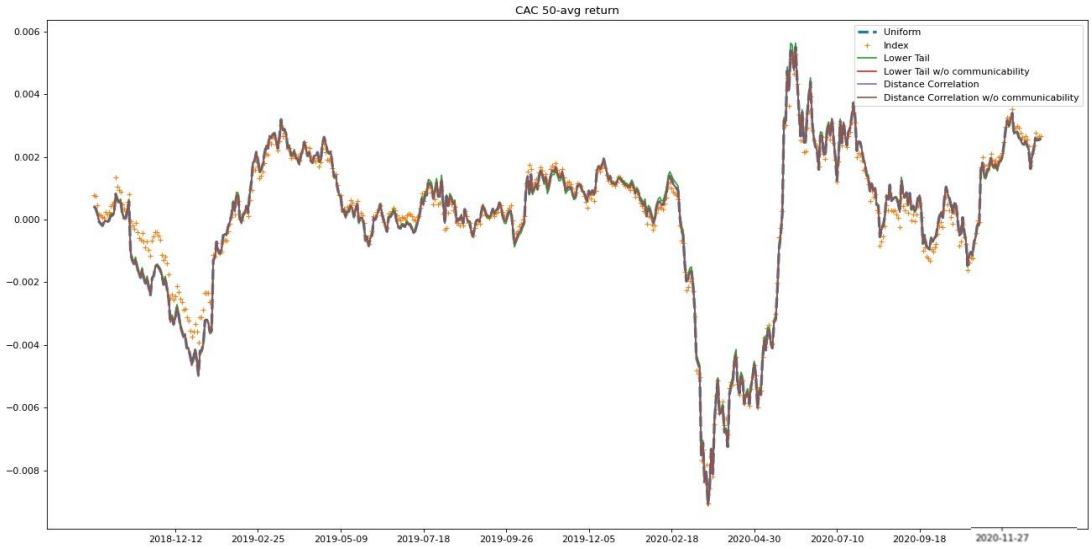
AVG


CAC capital
1e6
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
1.1
1.0
0.9
0.8
0.7
2018-01-02
2018-04-27
2018-12-12
2019-02-25
2019-05-09
2019-07-18
2019-09-26
2019-12-05
2020-02-18
2020-04-30
2020-07-10
2020-09-18
2020-11-27


CAC 50-avg return
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.008
2018-12-12
2019-02-25
2019-05-09
2019-07-18
2019-09-26
2019-12-05
2020-02-18
2020-04-30
2020-07-10
2020-09-18
2020-11-27

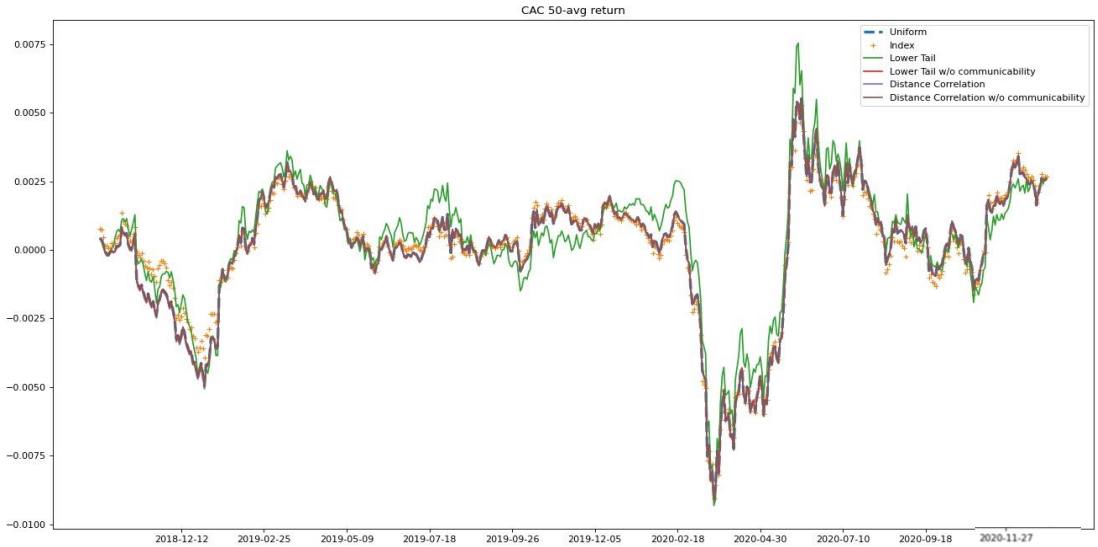
EXP


CAC capital
1e6
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
1.1
1.0
0.9
0.8
0.7
2018-01-02
2018-04-27
2018-12-12
2019-02-25
2019-05-09
2019-07-18
2019-09-26
2019-12-05
2020-02-18
2020-04-30
2020-07-10
2020-09-18
2020-11-27


CAC 50-avg return
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
0.006
0.004
0.002
0.000
−0.002
−0.004
−0.006
−0.008
2018-12-12
2019-02-25
2019-05-09
2019-07-18
2019-09-26
2019-12-05
2020-02-18
2020-04-30
2020-07-10
2020-09-18
2020-11-27

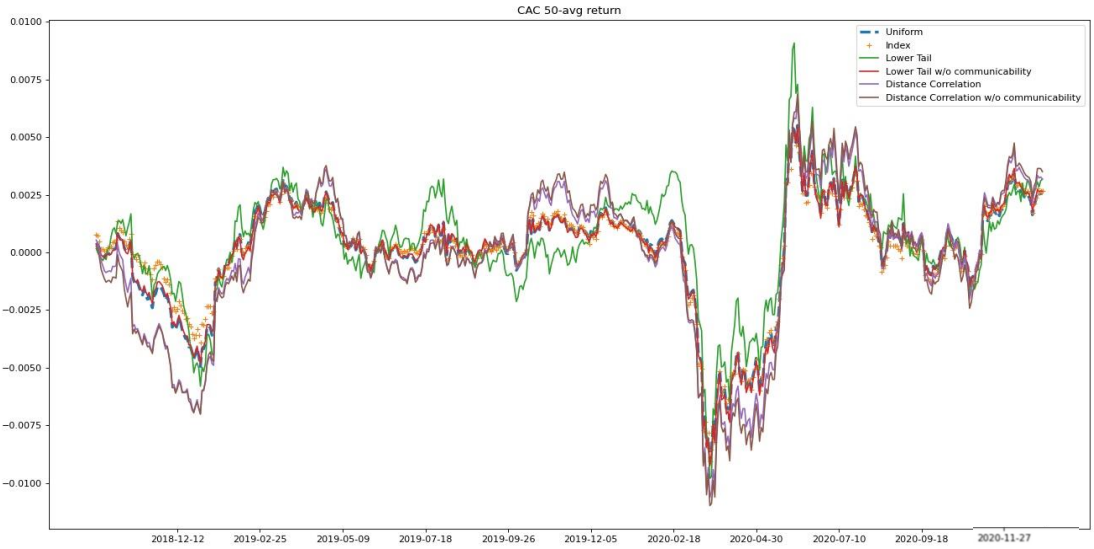
LN


CAC capital
1e6
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
1.2
1.1
1.0
0.9
0.8
0.7
2018-01-02
2018-04-27
2018-12-12
2019-02-25
2019-05-09
2019-07-18
2019-09-26
2019-12-05
2020-02-18
2020-04-30
2020-07-10
2020-09-18
2020-11-27


CAC 50-avg return
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
0.0075
0.0050
0.0025
0.0000
−0.0025
−0.0050
−0.0075
−0.0100
2018-12-12
2019-02-25
2019-05-09
2019-07-18
2019-09-26
2019-12-05
2020-02-18
2020-04-30
2020-07-10
2020-09-18
2020-11-27

TT


CAC capital
1e6
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
2018-01-02
2018-04-27
2018-12-12
2019-02-25
2019-05-09
2019-07-18
2019-09-26
2019-12-05
2020-02-18
2020-04-30
2020-07-10
2020-09-18
2020-11-27


CAC 50-avg return
Uniform
Index
Lower Tail
Lower Tail w/o communicability
Distance Correlation
Distance Correlation w/o communicability
0.0100
0.0075
0.0050
0.0025
0.0000
−0.0025
−0.0050
−0.0075
−0.0100
2018-12-12
2019-02-25
2019-05-09
2019-07-18
2019-09-26
2019-12-05
2020-02-18
2020-04-30
2020-07-10
2020-09-18
2020-11-27

Chapter 7. Discussion

From the Table 1 in the Results section, we can see that for every index and almost every allocation function, the lower-tail dependence associated with communicability centrality betweenness shows better performance than the other solutions. Even the uniform allocation and an index pegged portfolio do not overperform this method. The two exceptions are concerning the CAC and Kospi indexes. For CAC, the index allocation overperform the AVG allocation and has a close result to the EXP allocation. For Kospi, despite a slightly better average result, the performance is very close to the uniform allocation. This is due to the fact that the pruning threshold to preserve the connexe structure of the graph is extremely low compared to other markets. This may be due to an overall strong correlation of the underlying assets of Kospi.

By observing the allocation bar graphs, we can see that the different allocations functions provide different approach to the portfolio constitution: AVG and EXP provide a highly diversified portfolio with a share of every stock of the index and a slightly higher share attributed to assets with low centrality. On the other hand, LN and TT single out specific stock that can be considered as likely to outperform the market during a crisis. This translates by a higher return but a riskier portfolio as some stock can reach a 40% share of the asset bag.

Note: while the thresholds used to prune the correlation networks are made to preserve a connexe structure using lower tail dependence as edges weights, the thresholds for networks using distance correlation are far higher. However, a stronger pruning does not yield better performances during back testing, hence the minimum weight used for edge pruning on distance correlation is the same as the one for lower tail dependence.

Chapter 8. Conclusion

The experiment results show good performances as well as a good adaptability of the asset correlation network strategy. Overall, the communicability betweenness centrality using lower tail dependence as asset correlation proves to be the best method for the asset correlation network strategy. An interesting point is the fact that lowering the threshold for graph pruning draw the allocation closer to the uniform asset allocation, hence reducing the over-exposition to certain assets. This can be used to adapt the strategy to the risk tolerance of the investor or to define more precisely the strategy desired: this method can be used to build a diverse portfolio with little correlation with the market or it can be diverted to identify stocks with high potential during bear markets or high volatility periods.

Figures

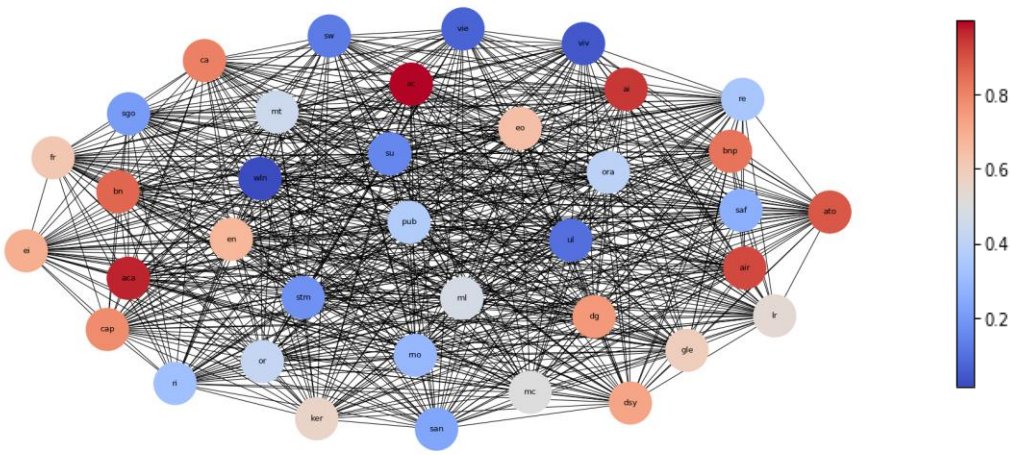


Figure 1.a) CAC Asset correlation network displaying centrality values

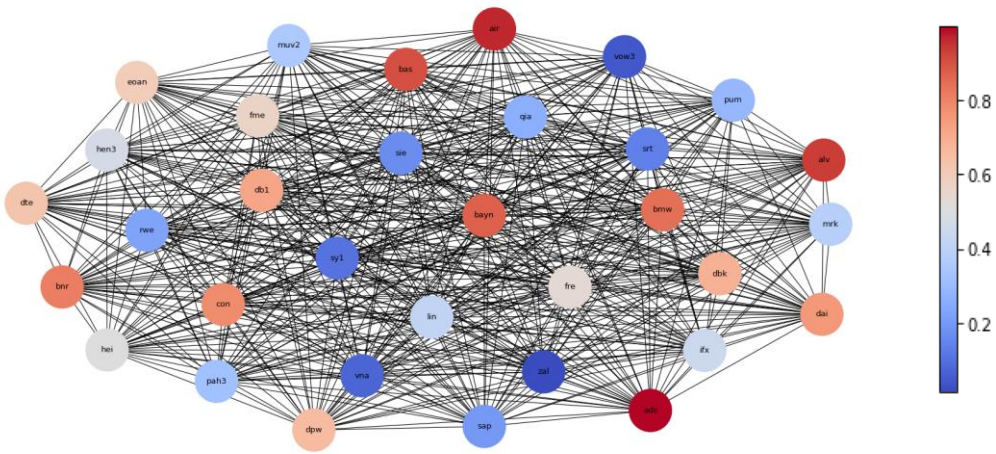


Figure 1.b) DAX Asset correlation network displaying centrality values

Figure 2.a) DAX correlation dendrogram using lowertail dependence as correlation factor

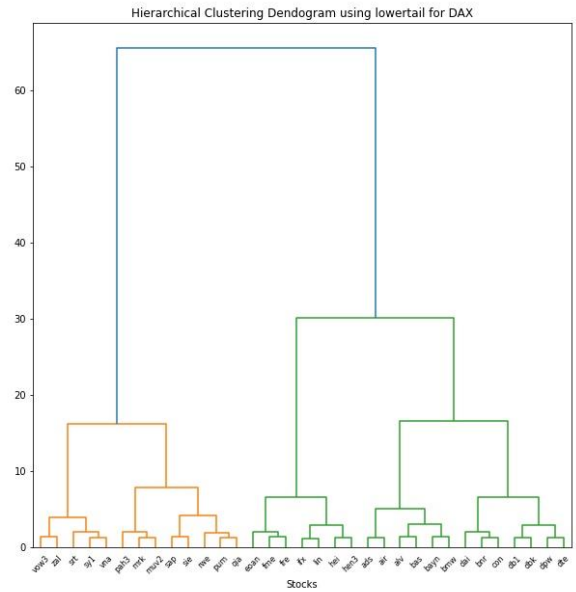


Figure 2.b) DAX correlation dendrogram using distance correlation as correlation factor

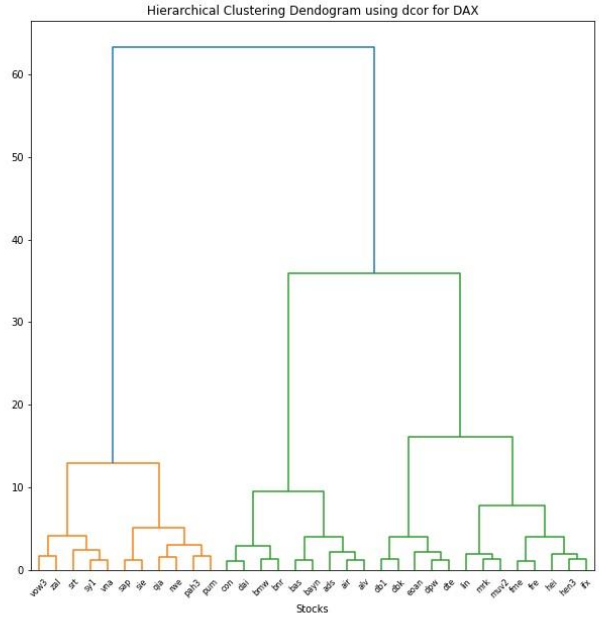


Figure 2.c) CAC correlation dendrogram using lowertail dependence as correlation factor

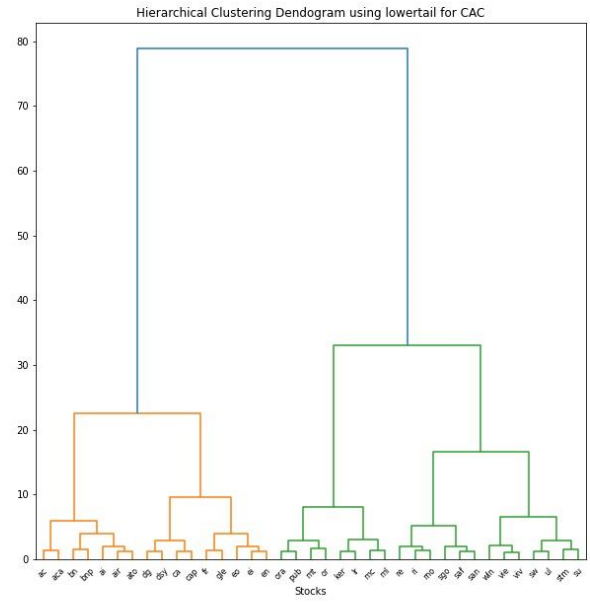


Figure 2.c) CAC correlation dendrogram using distance correlation as correlation factor

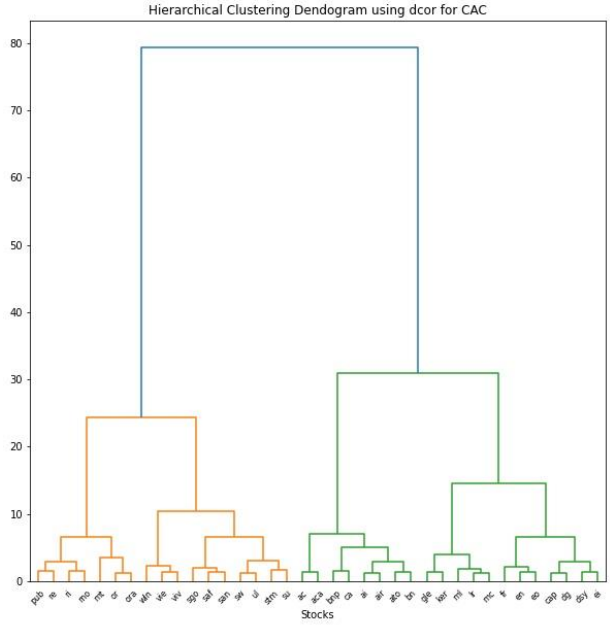


Figure 3.a) Copulas coefficient inference on DAX

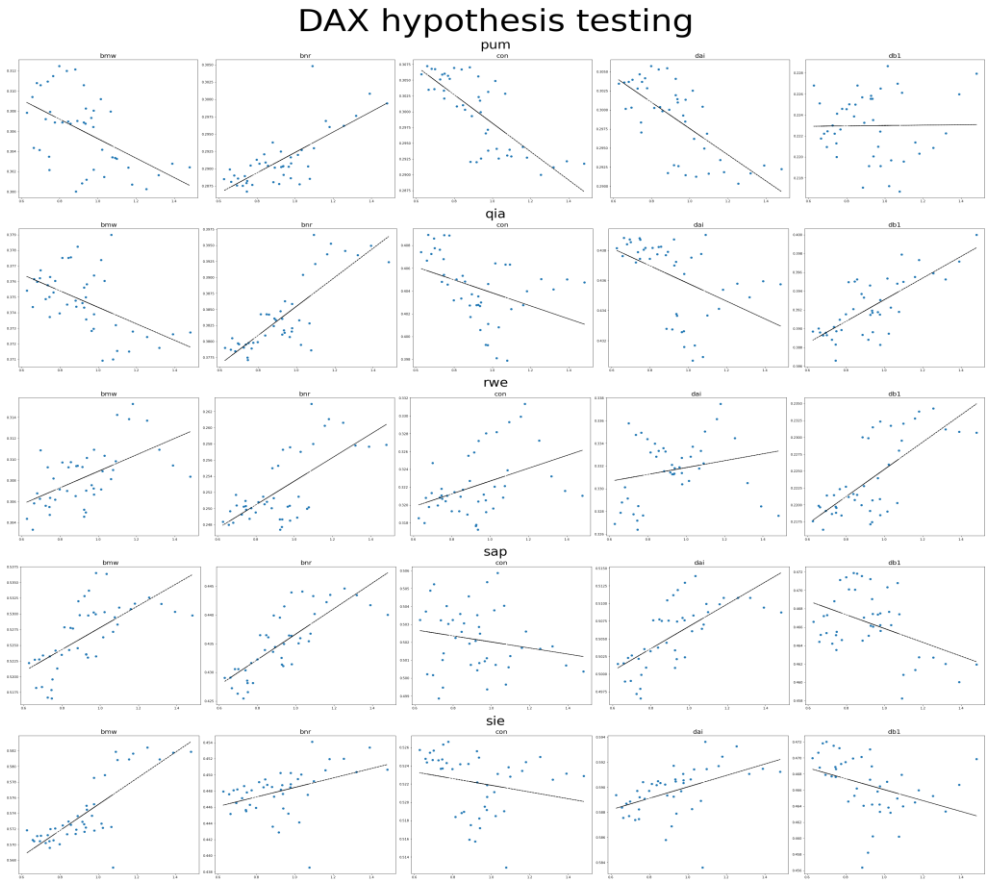
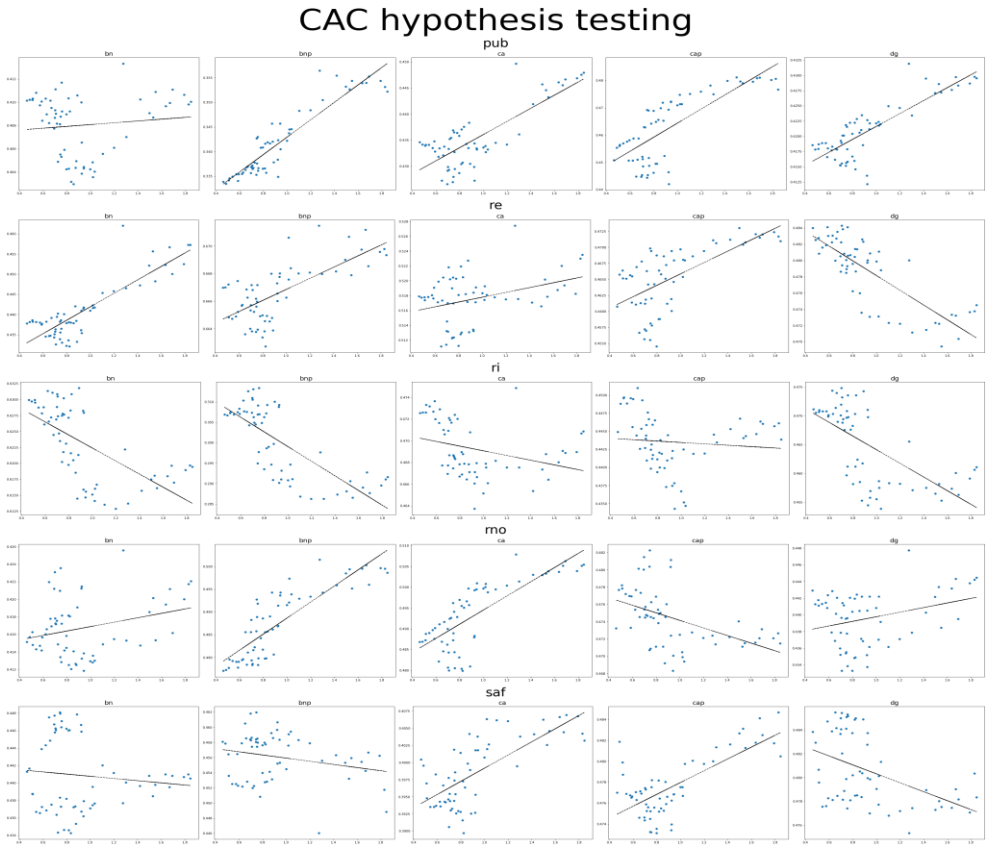


Figure 3.b) Copulas coefficient inference on CAC



Bibliography

[1] Community structure in social and biological networks, Michelle Girvan & M. E. J. Newman

[2] Finding and evaluating community structure in networks, Michelle Girvan & M. E. J. Newman

[3] Modularity and community structure in networks, M. E. J. Newman

[4] Communicability in complex networks, Ernesto Estrada & Naomichi Hatano

[5] Communicability Betweenness in Complex Networks, Ernesto Estrada, Desmond J. Higham & Naomichi Hatano

[6] A Tail Dependence−based dissimilarity measure for financial time series clustering, Giovanni De Luca and Paola Zuccolotto

[7] Dynamic clustering of financial assets, Giovanni De Luca and Paola Zuccolotto

Abstract

포트폴리오 다양화는 강력한 투자전략 수립에 있어 주요 관심사다. 자산 상관관계는 자산 할당을 결정하는 주요 지표다. 상관관계를 평가하는 방법은 자산을 교점으로, 쌍방향 상관관계는 선으로 나타내는 그래프를 통해 시장을 표현하는 것이다. 강한 상관관계를 띄는 집단을 확인하거나 한 자산이 국제시장과 얼마나 상관관계가 있는지 평가함으로써 탄력적인 포트폴리오를 구축할 수 있다.

안타깝게도, 많은 실제 시스템과같이, 최단경로를 기반으로 하는 커뮤니티 탐지를 위한 일반적인 접근은 실제 조건을 설명하지 않는다. 본 연구는 두 영역에 기초하여 그래프 구축과 좋은 포트폴리오 할당에 필요한 상관관계 계산 방법에 대한 건설적인 통찰력을 제공하고자 한다 1. 그래프의 전파성과 중심성 척도 그리고 2. 자산 상관관계 평가를 위한 낮은 꼬리 의존성.